

CREDIT AND FINANCIAL CONTAGION: A POINT PROCESS APPROACH

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Sai Man Simon Kwok

August 2012

© 2012 Sai Man Simon Kwok
ALL RIGHTS RESERVED

CREDIT AND FINANCIAL CONTAGION: A POINT PROCESS APPROACH

Sai Man Simon Kwok, Ph.D.

Cornell University 2012

This dissertation presents a set of econometric tools to uncover the mechanism of credit and financial contagion. First, a nonparametric Granger causality test for continuous time point process data is proposed. The test delivers inference results that are robust to model misspecifications. Applying the test to the point process data of Chapter 11 filings by U.S. corporations and negative shocks of major stock indices, the dissertation provides evidence for credit contagion across different sectors of the economy, as well as financial contagion across international stock markets. Second, a diagnostic checking procedure for parametric multivariate point process models is studied. The methodology equips empirical researchers with a portmanteau test in the crucial step of model validation after estimating a proposed parametric model.

BIOGRAPHICAL SKETCH

Simon Kwok had been a graduate student at the Economics department of Cornell University since August 2006. Before going to Cornell University, he finished the Master in Philosophy in Statistics under the supervision of Professor Wai Keung Li in July 2006 and the Bachelor of Science in Actuarial Science in May 2004, both with the Department of Statistics and Actuarial Science at the University of Hong Kong. His main research interests are econometrics, finance and time series analysis.

To my parents and Carol.

ACKNOWLEDGEMENTS

I am greatly indebted to my advisor, Professor Yongmiao Hong, who provided me with invaluable guidance and constant encouragement. My heartfelt thank also goes to my special committee members, Professor Nicholas Kiefer, Professor Robert Jarrow and Professor Robert Strawderman. Their useful feedback and sharp criticisms have been crucial in shaping my final thesis. I am also thankful for the feedback from the audience of my presentations in conferences and seminars over the past year, and my fellow colleagues would be remembered for their support and tolerance over my numerous presentation practices. Last but not least, I have to thank my wife, Carol Xu, for staying by my side throughout the difficult periods and for showing her understanding, without which this dissertation would be non-existent.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
1 Introduction	1
2 A Nonparametric Test of Granger Causality in Continuous Time	4
2.1 Introduction	4
2.1.1 The Need for Continuous Time Causality Test	8
2.1.2 The Need for Nonparametric Causality Test	11
2.1.3 Point Processes in High Frequency Finance	13
2.1.4 Point Processes in Counterparty Risk Modeling	16
2.1.5 Test of Dependence between two stochastic processes	18
2.2 Bivariate Point Process	19
2.3 Granger Causality	28
2.4 The statistic	33
2.4.1 Nonparametric cross-covariance estimator	33
2.4.2 The statistic as \mathcal{L}^2 norm	35
2.4.3 Weighting function	36
2.4.4 Conditional intensity estimator	37
2.4.5 Computation of $\hat{\gamma}_H(\ell)$	39
2.4.6 Consistency of conditional intensity estimator	39
2.4.7 Simplified Statistic	44
2.5 Asymptotic Theory	48
2.5.1 Asymptotic Normality under the Null	48
2.5.2 Effect of Estimation	49
2.5.3 Asymptotic Local Power	51
2.6 Bandwidth Choices	52
2.6.1 Case 1: $B \ll H \ll T$	53
2.6.2 Case 2: $H \ll B \ll T$	53
2.6.3 Optimal Bandwidths	54
2.7 Simulations	55
2.7.1 Size and Power of Q	55
2.7.2 Size and Power of Q^s	56
2.8 Applications	60
2.8.1 Trades and Quotes	60
2.8.2 Credit Contagion	64
2.8.3 International Financial Contagion	67
2.9 Conclusion	70

3	Credit Contagion from Wall Street to Main Street: An Empirical Study of US Corporate Bankruptcies	72
3.1	introduction	72
3.2	VAR(1) model	80
3.2.1	Spurious causality	81
3.2.2	Spurious non-causality	83
3.3	Bivariate ACI model	83
3.4	Nonparametric Granger causality test	85
3.5	Empirical results	85
3.5.1	Bankruptcy Contagion: upstream vs downstream	86
3.5.2	Bankruptcy Contagion: Wall Street vs Main Street	87
4	Diagnostic Checks for Multivariate Parametric Intensity Models	89
4.1	Introduction	89
4.2	Parametric multivariate point process models	92
4.3	Random Time Change and Characterization of Multivariate In-	
	tensity Models	93
4.3.1	Generalized Residual Autocorrelations Test	99
4.4	Simulations	103
5	Conclusion	109
A	Chapter 2 appendix	111
A.1	List of Assumptions	111
A.2	Figures	113
A.3	Tables	126
A.4	Proof of Theorem 5	135
A.5	Proof of Theorem 6	136
A.6	Proof of Theorem 7	138
A.7	Proof of (2.21)	141
A.8	Proof of Theorem 8	143
A.8.1	Asymptotic Mean of Q	143
A.8.2	Asymptotic Variance of Q Under the Null	145
A.8.3	Asymptotic normality of \tilde{Q}	154
A.9	Proof of Theorem 10	164
A.10	Proof of Theorem 9	165
A.11	Proof of Corollary 11	167
A.12	Proof of Corollary 12	169
A.13	Summary of Jarrow and Yu (2001) Model	170
A.14	Q Test for Non-stationary Point Processes	172
B	Chapter 3 appendix	175

C Chapter 4 appendix	188
C.1 Proof of Theorem 19	188
Bibliography	194

LIST OF TABLES

A.1	The asymptotic mechanisms of the two schemes.	126
A.2	Significant day counts (out of 41 days) of PG, 9:45am – 10:15am.	127
A.3	Significant day counts (out of 41 days) of PG, 11:45am – 12:45pm.	127
A.4	Significant day counts (out of 41 days) of PG, 3:30pm – 4:00pm.	128
A.5	Significant day counts (out of 41 days) of PG over various trading hours of a day.	129
A.6	Significant day counts (out of 41 days) of GM over various trading hours of a day.	129
A.7	Q tests on bankruptcy data, Sep96 – Jul03.	130
A.8	Q test on bankruptcy data, Aug03 – Aug07.	130
A.9	Q tests on bankruptcy data, Sep07 – Jun10.	131
A.10	Q^s test on bankruptcy data, September 1996 – June 2010.	132
A.11	Trading hours, Greenwich mean time and start dates of the sampling periods of major stock indices.	132
A.12	Q^s test applied to extreme negative shocks of DJI and HSI.	133
A.13	Q^s test applied to extreme negative shocks of DJI and NIK.	133
A.14	Q^s test applied to extreme negative shocks of DJI and FTSE.	133
A.15	Q^s test applied to extreme negative shocks of DJI and AOL.	134
B.1	Estimated bivariate ACI(1,1) models on point process data of bankruptcy filings over recessions and crises.	176
B.2	Estimated bivariate ACI(1,1) models on point process data of bankruptcy filings over non-recession periods.	177
B.3	Estimated VAR(1) model on the time series of bankruptcy counts over Dec 2007 - June 2009 with varying sampling frequencies.	178
B.4	Estimated VAR(1) model on the time series of bankruptcy counts over March 2001 - November 2002 with varying sampling frequencies.	179
B.5	Estimated VAR(1) model on the time series of bankruptcy counts over July 1990 - March 1991 with varying sampling frequencies.	180
B.6	Estimated VAR(1) model on the time series of bankruptcy counts over April 1990 - March 1991 with varying sampling frequencies.	181
B.7	Estimated VAR(1) model on the time series of bankruptcy counts over December 2002 - November, 2007 with varying sampling frequencies.	182
B.8	Estimated VAR(1) model on the time series of bankruptcy counts over July, 2009 - November, 2011 with varying sampling frequencies.	183

LIST OF FIGURES

4.1	The empirical sizes of various portmanteau tests. DGP: ACI(1,1), df= $M, n = 100$	105
4.2	The empirical sizes of various portmanteau tests. DGP: ACI(1,1), df= $M, n = 400$	105
4.3	The empirical sizes of various portmanteau tests. DGP: ACI(1,1), df= $M - 2, n = 100$	105
4.4	The empirical sizes of various portmanteau tests. DGP: ACI(1,1), df= $M - 2, n = 400$	106
4.5	The empirical sizes of various portmanteau tests. DGP: Cox model, df= $M, n = 100$	106
4.6	The empirical sizes of various portmanteau tests. DGP: Cox model, df= $M, n = 400$	106
4.7	The empirical sizes of Box-Pierce and Ljung-Box tests. DGP: bi- variate ACI(1,1) model, df= $M, n = 100$	107
4.8	The empirical sizes of Box-Pierce and Ljung-Box tests. DGP: bi- variate ACI(1,1) model, df= $M, n = 400$	107
4.9	The empirical power of Box-Pierce and Ljung-Box tests. DGP: bivariate ACI(1,1) model, df= $M, n = 100$	108
4.10	The empirical power of Box-Pierce and Ljung-Box tests. DGP: bivariate ACI(1,1) model, df= $M, n = 400$	108
A.1	The statistic Q aggregates the squared contributions of residual products $d\hat{\epsilon}_s^a d\hat{\epsilon}_t^b$ for all $s < t$. The lines join all pairs of type a and type b events (shocks) at their event times (τ_i^a, τ_j^b) for all $\tau_i^a < \tau_j^b$	113
A.2	Size experiment of Q test, bivariate Poisson process.	114
A.3	Size and power experiment of Q test, bivariate exponential Hawkes process.	115
A.4	Size experiment 1 of Q^s test.	116
A.5	Size experiment 2 of Q^s test.	117
A.6	Size experiment 3 of Q^s test.	118
A.7	Size experiment 4 of Q^s test.	119
A.8	Size experiment 5 of Q^s test.	120
A.9	Power experiment of Q^s test.	121
A.10	Size experiment R-S1 of Q^s test.	122
A.11	Size experiment R-S2 of Q^s test.	123
A.12	Power experiment R-P1 of Q^s test.	124
A.13	Power experiment R-P2 of Q^s test.	125
A.14	Histogram of bankruptcies of U.S. firms, 1980–2010.	125
A.15	Raw counts of bankruptcies in manufacturing and financial re- lated sectors.	126
B.1	Q^s test between upstream and downstream of a supply chain during crises and recessions.	184

B.2	Q^s test between upstream and downstream of a supply chain during non-recession periods.	185
B.3	Q^s test between Wall Street and Main Street during crises and recessions.	186
B.4	Q^s test between Wall Street and Main Street during non-recession periods.	187

CHAPTER 1

INTRODUCTION

Credit and financial crises refer to the breakdown of the credit and financial markets due to their fragility. Rare and random as they are, the impact to the economy is often devastating and widespread once they start. However, it is often difficult to identify a single cause for a crisis. *Credit contagion* results when credit risk, taking the form of such credit events as individual defaults and bankruptcies, spreads across different sectors of the economy in an uncontrollable manner. Similarly, *financial contagion* occurs when a significant negative shock to a financial market creates panic that transmits across different sectors and financial markets. The distinctive characteristics include the the infectious nature and the severity of the impact.

From the statistical point of view, the sequences of credit events and negative shocks are random and irregular-spaced events over time. They are *point processes* that evolve in continuous time and display certain stylized facts such as *self-excitation* (or clustering of events of the same type) and *mutual-excitation* (or feedback from a cluster of events of one type to a cluster of events of another type). These stylized facts can be measured and analyzed in a parametric or a nonparametric set-up. To this end, I am going to consider a set of tools that are useful for the analysis of credit and financial crises.

In Chapter 2, I consider a nonparametric Granger causality test for continuous time point process data. Unlike popular Granger causality tests with strong parametric assumptions on discrete time series, the test applies directly to strictly increasing raw event time sequences sampled from a bivariate temporal point process satisfying mild stationarity and moment conditions, thus

eliminating the sensitivity of the test to model assumptions and data sampling frequency. Taking the form of an L^2 -norm, the test statistic delivers a consistent test against all alternatives with pairwise causal feedback from one component process to another, and can simultaneously detect multiple causal relationship over variable ranges up to the sample length. The test enjoys asymptotic normality under the null of no Granger causality and exhibits reasonable empirical size and power performance. Its usefulness is illustrated in three applications. In the first application on the study of market microstructure hypotheses, the test confirms the existence of a significant causal relationship from the dynamics of trades to quote revisions in high frequency financial datasets. The next application on credit contagion reveals that corporate bankruptcies in financial related sectors tend to Granger-cause those in manufacturing related sectors during crises and recessions. Lastly, the test is applied to study the extent to which an extreme negative shock of a major stock index transmits across international financial markets. The test confirms the presence of contagion, with US and European stock indices being the major sources of contagion.

In Chapter 3, I investigate empirically the contagious relationship of US corporate bankruptcies from the upstream and downstream of a supply chain. The data span from 1980 to 2011. The observed lead-lag pattern can be thought of as the aggregated result of a complex web of mutual dependence between counterparties in the two sectors. Applying a nonparametric Granger causality test in continuous time, I provided empirical evidence of credit contagion spreading from the bottom to the top of a typical supply chain around recession periods. It is believed that such empirical results would shed light on investors who maintain portfolios consisting of credit derivatives, and help the government to tailor its monetary and fiscal policies to avoid a major breakdown of the economy.

In Chapter 4, I turn to the problem of diagnostic checking multivariate parametric intensity models, which are useful for the empirical study of high frequency finance and credit risk. Since they were often proposed for a particular application without much theoretical justification, specification tests are crucial for ensuring model adequacy and the validity of statistical inference. However, unlike traditional discrete time series models which are equipped with well-studied portmanteau tests, there are neither theoretical results nor empirical performance analyses of such tests applied to an estimated multivariate intensity model with parameter uncertainty. This chapter aims at filling this gap by deriving a large sample distribution for the generalized residual autocorrelations and proposing a portmanteau test that checks for model adequacy. The test procedures are theoretically justified for a wide class of multivariate parametric recurrent-event intensity models.

CHAPTER 2

A NONPARAMETRIC TEST OF GRANGER CAUSALITY IN CONTINUOUS TIME

2.1 Introduction

The concept of Granger causality was first introduced to econometrics in the ground-breaking work of Granger (1969) and Sims (1972). Since then it has generated an extensive line of research and quickly became a standard topic in econometrics and time series analysis textbooks. The idea is straightforward: a process X_t does not strongly (weakly) Granger cause another process Y_t if, at all time t , the conditional distribution (expectation) of Y_t given its own history is the same as that given the histories of both X_t and Y_t almost surely. Intuitively, it means that the history of process X_t does not affect the prediction of process Y_t .

Granger causality tests are abundant in economics and finance. Instead of giving a general overview on Granger causality tests, I will focus on some of the shortfalls of popular causality tests. Currently, most Granger causality tests in empirical applications rely on parametric assumptions, most notably the discrete time vector autoregressive (VAR) models. Although it is convenient to base the tests on discrete time parametric models, there are a couple of issues that can potentially invalidate this approach:

(1) Model uncertainty. If the data generating process (DGP) is far from the parametric model, the econometrician will run the risk of model misspecification. The conclusion of a Granger causality test drawn from a wrong model can be misleading. A series of studies attempts to reduce the effect of model un-

certainty by relaxing or eliminating the reliance on strong parametric assumptions.¹

(2) Sampling frequency uncertainty. Existing tests of Granger causality in discrete time often assume that the time difference between consecutive observations is constant and prespecified. However, it is important to realize that the conclusion of a Granger causality test can be sensitive to the sampling frequency of the time series. As implied by the results of Sims (1971) and argued by Engle and Liu (1972), the test would potentially be biased if we estimated a discretized time series model with temporally aggregated data which are from a continuous time DGP (see section 2.1.1).

To address the above shortcomings, I consider a nonparametric Granger causality test in continuous time. The test is independent of any parametric model and thus the first problem is eliminated. Unlike discrete time Granger causality tests, the test applies to data sampled in continuous time - the highest sampling frequency possible - and can simultaneously and consistently detect causal relationships of various durations spanning up to the sample length. The DGP is taken to be a pure-jump process known as *bivariate temporal point process*.

A *temporal point process* is one of the simplest kinds of stochastic process and is the central object of this chapter. It is a pure-jump process consisting of a sequence of *events* represented by jumps that occur over a continuum, and the observations are event occurrence times (called *event times*).² Apart from

¹One line of research extends the test to nonlinear Granger causality test. To relax the strong linear assumption in VAR models, Hiemstra and Jones (1994) developed a nonparametric Granger causality tests on discrete time series without imposing any parametric structures on the DGP except some mild ones such as stationarity and Markovian dynamics. In the application of their test, they found that volume Granger causes stock return.

²The trajectory of a *counting process*, an equivalent representation constructed from point process observations, is a stepwise increasing and right-continuous function with a jump at each event time. An important example is the Poisson process in which events occur independently

their simplicity, point processes are indispensable building blocks of other more complicated stochastic processes (e.g. Lévy processes, subordinated diffusion processes). In this chapter, I study the testing of Granger causality in the context of a *simple*³ *bivariate point process*, which consists of a strictly monotonic sequence of event times originated from two event types with possible interactions among them. The problem of testing Granger causality consistently and nonparametrically in a continuous time set-up for a simple bivariate point process is non-trivial: all interactive relationship of event times over the continuum of the sample period needs to be summarized in a test statistic, and continuous time martingale theory is necessary to analyze its asymptotic properties. It is hoped that the results reported in this chapter will shed light on a similar test for more general types of stochastic processes.

To examine the causal relation between two point processes, I first construct event counts (as a function of time) of the two types of events from the observed event times. The functions of event counts, also known as *counting processes*, are monotone increasing functions by construction. To remove the increasing trends, I consider the differentials of the two counting processes. After subtracting their respective conditional means (estimated nonparametrically), I obtain the *innovation processes* that contain the surprise components of the point processes. It is possible to check, from the cross-covariance between the innovation processes, if there is a significant feedback from one counting process to another. As detailed in section 2.2, such a feedback relationship is linked to the Granger causality concept that was defined for general continuous time processes (including counting processes as a particular case) in the extant literature. More surprisingly, if the raw event times are strictly monotonic, then *all pairwise cross-*

of each other.

³The *simple* property will be formally defined in assumption (A1) in section 2.2.

dependence can be sufficiently captured by the *cross-covariance* between the innovation processes. This insight comes from the Bernoulli nature of the jump increments of the associated counting processes, and will greatly facilitate the development and implementation of the test.

The chapter is organized as follows. Empirical applications of point processes are described in sections 2.1.3 and 2.1.4. The relevant concepts and properties of a simple bivariate point process is introduced in section 2.2, while the concept of Granger causality is discussed and adapted to the context of point processes in section 2.3. The test statistic is constructed in section 2.4 as a weighted integral of the squared cross-covariance between the innovation processes, and the key results on its asymptotic behaviors are presented in section 2.5. Variants of the test statistic under different bandwidth choices are discussed in section 2.6. In the simulation experiments in section 2.7, I show that the nonparametric test has reasonable size performance under the null hypothesis of no Granger causality and nontrivial power against different alternatives. In section 2.8, I demonstrate the usefulness of the nonparametric Granger causality test in a series of three empirical applications. In the first application on the study of market microstructure hypotheses (section 2.8.1), we see that the test confirms the existence of a significant causal relationship from trades to quote revisions in high frequency financial datasets. Next, I turn to the application in credit contagion (section 2.8.2) and provide the first empirical evidence that bankruptcies in financial-related sectors tend to Granger-cause those in manufacturing-related sectors during crises and recessions. In the last application on international financial contagion (section 2.8.3), I examine the extent to which an extreme negative shock of a major stock index is transmitted across international financial markets. The test reveals the presence of financial contagion, with U.S. and

European stock indices being the sources of contagion. Finally, the chapter concludes in section 2.9. Proofs and derivations are collected in Appendix A.

2.1.1 The Need for Continuous Time Causality Test

The original definition of Granger causality is not only confined to discrete time series but also applicable to continuous time stochastic processes. However, an overwhelming majority of research work on Granger causality tests, be it theoretical or empirical, has focused on a discrete time framework. One key reason for this is the limited availability of (near) continuous time data. However, with much improved computing power and storage capacity, economic and financial data sampled at increasingly high frequencies have become more accessible.⁴ This calls for more sophisticated techniques for analyzing these datasets. To this end, continuous time models provide a better approximation to frequently observed data than discrete time series models with very short time lags and many time steps. Indeed, even though the data are observed and recorded in discrete time, it is sometimes more natural to think of the DGP as evolving in continuous time, because economic agents do not necessarily make decisions at the same time when the data are sampled. The advantages of continuous time analyses are more pronounced when the observations are sampled (or available) at random time points. Imposing a fixed discrete time grid on highly irregularly spaced time data may lead to too many observations in frequently sampled periods and/or excessive null intervals with no observations in sparsely sampled periods.⁵

⁴For example, trade and quote data now include records of trade and quote timestamps in unit of milliseconds.

⁵Continuous time models are more parsimonious for modeling high frequency observations and are more capable of endogenizing irregular and possibly random observation times. See,

Furthermore, discretization in time dimension can result in the loss of time point data and *spurious (non)causality*. The latter problem often arises when “the finite time delay between cause and effect is small compared to the time interval over which data is collected”, as pointed out by Granger (1988, p.205). A Granger causality test applied to coarsely sampled data can deliver very misleading results: while the DGP implies a unidirectional causality from process X_t to process Y_t , the test may indicate (i) a significant *bidirectional causality* between X_t and Y_t , or (ii) insignificant causality between X_t and Y_t in either one or both directions.⁶ The intuitive reason is that the causality of the discretized series is the aggregate result of the causal effects in each sampling intervals, amplified or diminished by the autocorrelations of the marginal processes. The severity of these problems depends on prespecified sampling intervals: the wider they are relative to the *causal durations* (the actual time durations in which causality effect transmits), the more serious the problems.⁷ With increasingly accessible high frequency and irregularly spaced data, it is necessary to develop theories and techniques tailored to the continuous time framework to uncover any interactive patterns between stochastic processes. Analyzing (near) continuous time data with inappropriate discrete time techniques is often the culprit of misleading conclusions.

To remedy the above problems, there have been theoretical attempts to extend discrete time causality analyses to fully continuous time settings. For

for instance, Duffie and Glynn (2004), Aït-Sahalia and Mykland (2003), Li, Mykland, Renault, Zhang, Zheng (2010).

⁶Sims (1971) provided the first theoretical explanation in the context of distributed lag model (a continuous time analog of autoregressive model). See also Geweke (1978), Christiano and Eichenbaum (1987), Marcet (1991) and, for a more recent survey, McCrorie and Chambers (2006).

⁷For instance, suppose the DGP implies a causal relationship between two economic variables which typically lasts for less than a month. A Granger causality test applied to the two variables sampled weekly can potentially reveal a significant causal relationship, but the test result may turn insignificant if applied to the same variables sampled monthly.

example, Florens and Fougere (1996) examined the relationship between different definitions of Granger non-causality for general continuous time models. Comte and Renault (1996) studied a continuous time version of ARMA model and provided conditions on parameters that characterize when there is no Granger causality, while Renault, Sekkat and Szafarz (1998) gave corresponding characterizations for parametric Markov processes. All of the above work, however, did not elaborate further on the implementation of the tests, let alone any formal test statistic and empirical applications.

Due to a lack of continuous time testing tools for high-frequency data, practitioners generally rely on parametric discrete time series methodology or multivariate parametric point process models. Traditionally, time series econometricians have little choice but to adhere to a fixed sampling frequency of the available dataset, even though they have been making an effort to obtain more accurate inference by using the highest sampling frequency that the data allow (Engle, 2000). The need to relax the rigid sampling frequency is addressed by the literature on mixed frequency time series analyses.⁸ On the other hand, inferring causal relationships from parametric point process models may address some of these problems as this approach respects the irregular nature of event times.

It is important to reiterate that correct inference about the directions of Granger causality stems from an appropriate choice of sampling grid. The actual causal durations, however, are often unknown or even random over time (as is the case for high-frequency financial data). In light of this reality, it is more

⁸Ghysels (2012) extends the previous mixed frequency regression to VAR models with a mixture of two sampling frequencies. Chiu, Eraker, Foerster, Kim and Seoane (2011) proposed a Bayesian mixed frequency VAR models which are suitable for irregularly sampled data. This kind of models has power for DGPs in which Granger causality acts over varying horizons.

appealing to carry out Granger causality tests on continuous time processes in a way that is independent of the choice of sampling intervals and allows for simultaneous testing of causal relationships with variable ranges.

2.1.2 The Need for Nonparametric Causality Test

Often times, the modelers adopt a pragmatic approach when choosing a parametric model in order to match the model features to the observed stylized facts of the data. In the study of empirical market microstructure, there exist parametric bivariate point process models that explain trade and quote sequences. For example, Russell (1999) proposed the flexible multivariate autoregressive conditional intensity (MACI) model, in which the intensity takes a log-ARMA structure that explains the clustered nature of tick events (trades and quotes). More recently, Bowsher (2007) generalized the Hawkes model (gHawkes), formerly applied to clustered processes such as earthquakes and neural spikes, to accommodate intraday seasonality and interday dependence features of high frequency TAQ data. Even though structural models exist that predict the existence of Granger causality between trade and quote sequences, the functional forms of the intensity functions are hardly justified by economic theories. Apart from their lack of theoretical foundation, MACI and gHawkes models were often inadequate for explaining all the observed clustering in high frequency data, as evidenced by unsatisfactory goodness-of-fit test results. Model misspecification may potentially bias the result of causal inference. Hence, it would be ideal to have a nonparametric test that provides robust and model-free results on the causal dynamics of the data.

In this chapter, I pursue an alternative approach by considering a nonparametric test of Granger causality that does not rely on any parametric assumption and thus is free from the risk of model misspecification. Since I assume no parametric assumptions and only impose standard requirements on kernel functions and smoothing parameters, the conclusion of the test is expected to be more robust than existing techniques. In addition, the nonparametric test in this chapter can be regarded as a measure of the strength of Granger causality over different spans as the bandwidth of the weight function varies. Such "impulse response" profile is an indispensable tool in the quest for suitable parametric models.

More importantly, the conclusions from any statistical inference exercise are model specific and have to be interpreted with care. In other words, all interpretations from an estimated model are valid only under the assumption that the parametric model represents the true DGP. For example, in the credit risk literature, there has been ongoing discussion on whether the conditional independence model or the self-exciting clustering model provides a better description of the stylized facts of default data. This is certainly a valid goodness-of-fit problem from the statistical point of view, but it is dangerous to infer that the preferred model represents the true DGP. There may be more than one point process model that can generate the same dataset.⁹ The conclusion can entail substantial economic consequences: under a doubly stochastic model, credit contagion is believed to spread through information channels (Bayesian learning on common factors); while under a clustering model, credit contagion is transmitted through direct business links (i.e. counterparty risk exposure). The

⁹An example is provided by Barlett (1964), which showed that it is mathematically impossible to distinguish a linear doubly stochastic model and a clustering model with a Poisson parent process and one generation of offsprings (each of which is independently and identically distributed around each parent), as their characteristic functions are identical.

two families of DGPs are very different in both model forms and economic contents, but they can generate virtually indistinguishable data (Barlett, 1964). Without further assumptions, we are unable to differentiate the two schools of belief solely based on empirical analyses of Granger causality. It is precisely the untestability and non-uniqueness of model assumptions that necessitate a model-free way of uncovering the causal dynamics of a point process.

2.1.3 Point Processes in High Frequency Finance

Point process models are prevalent in modeling trade and quote tick sequences in high frequency finance. The theoretical motivation comes from the seminal work by Easley and O'hara (1992), who suggested that transaction time is endogenous to stock return dynamics and plays a crucial role in the formation of a dealer's belief in the fundamental stock price. Extending Glosten and Milgrom's (1985) static sequential trade model, their dynamic Bayesian model yields testable implications regarding the relation between trade frequency and the amount of information disseminated to the market, as reflected in the spread and bid/ask quotes set by dealers.

In one of the first empirical analyses, Hasbrouck (1991) applied a discrete vector autoregressive (VAR) model to examine the interaction between trades and quote revisions. Dufour and Engle (2000) extended Hasbrouck's work by considering time duration between consecutive trades as an additional regressor of quote change. They found a negative correlation between a trade-to-trade duration and the next trade-to-quote duration, thus confirming that trade intensity has an impact on the updating of beliefs on fundamental prices.¹⁰

¹⁰See Hasbrouck (2007, p.53) for more details.

Given the conjecture of Easley and O'hara (1992) and the empirical evidence of Dufour and Engle (2000), it is important to have a way to extract and model transaction time, which may contain valuable information about the dynamics of quote prices. To this end, Engle and Russell (1998) proposed the Autoregressive Conditional Duration (ACD) model, which became popular for modeling tick data in high frequency finance. It is well known that stock transactions on the tick level tend to cluster over time, and time durations between consecutive trades exhibit strong and persistent autocorrelations. The ACD model is capable of capturing these stylized facts by imposing an autoregressive structure on the time series of trade durations.¹¹

A problem with duration models is the lack of a natural multivariate extension due to the unsynchronized nature of trade and quote durations by construction (i.e. a trade duration always starts and ends in the middle of some other quote durations). At the time a trade occurs, the econometrician's information set would be updated to reflect the new trade arrival, but it is difficult to transmit the updated information to the dynamic equation for quote durations, because the current quote duration has not ended yet. The same difficulty arises when information from a new quote arrival needs to be transmitted in the opposite direction to the trade dynamics. Indeed, as argued by Granger (1988, p.206), the problem stems from the fact that durations are flow variables. As a result, it is impossible to identify clearly the causal direction between two flow variable sequences when the flow variables overlap one another in time dimension.¹² Nevertheless, there exist a number of methods that attempt to get

¹¹Existing applications of ACD model to trade and quote data are widespread, including (but not limited to) estimation of price volatility from tick data, testing of market microstructure hypotheses regarding spread and volume and intraday value-at-risk estimation. See Pacurar (2008) for a survey on ACD models.

¹²As another example, Renault and Werker (2011) tested for a causal relationship between quote durations and price volatility. They assume that tick-by-tick stock returns are sam-

around this problem, such as transforming the tick data to event counts over a prespecified time grid (Heinen and Rengifo, 2007) and redefining trade/quote durations in an asymmetric manner to avoid overlapping of durations (Engle and Lunde, 2003). They are not perfect solutions either.¹³

It is possible to mitigate the information transmission problem in a systematic manner, but this requires a change of viewpoint: we may characterize a multivariate point process from the point of view of *intensities* rather than duration sequences. The *intensity function* of a point process, which is better known as *hazard function* or *hazard rate* for more specific types of point processes in biostatistics, quantifies the event arrival rate at every time instant. Technically, it is the probability that at least one event occurs. While duration is a *flow* concept, event arrival rate is a *stock* concept and thus not susceptible to the information transmission problem. To specify a complete dynamic model for event times, it is necessary to introduce the concept of *conditional intensity function*: the conditional probability of having at least one event at the next instant given the history of the entire multivariate point process up to the present. The dynamics of different type events can be fully characterized by the corresponding conditional intensity functions. Russell (1999), Hautsch and Bauwens (2006), and Bowsher (2007) proposed some prominent examples of intensity models.¹⁴ The

pled from a continuous time Lévy process. Based on the moment conditions implied from the assumptions, they uncovered instantaneous causality from quote update dynamics to price volatility calculated from tick-by-tick returns. Similar criticism on Engle and Lunde (2003) applies to this work as well because trade durations over which volatility is computed overlap with quote durations.

¹³Information about durations is lost under the event count model of Heinen and Rengifo. Data loss problem occurs in the Engle and Lunde model when there are multiple consecutive quote revisions, as only the quote revision immediately after a trade is used. Moreover, the asymmetry of the Engle and Lunde model only allows the detection of trade-to-quote causality but not vice versa.

¹⁴Russell (1999) estimated a bivariate ACI model to uncover the causal relationship between transaction and limit order arrivals of FedEx from November 1990 to January 1991. With the gHawkes model, Bowsher (2007) provided empirical evidence of significant two-way Granger causality between trade arrivals and mid-quote updates of GM traded on the NYSE over a 40

objective is to infer the direction and strength of the lead-lag dependence among the marginal point processes from the proposed parametric model.

2.1.4 Point Processes in Counterparty Risk Modeling

The Granger causality test can be useful to test for the existence of counterparty risk in credit risk analysis. Counterparty risk was first analyzed in a bivariate reduced form model in Jarrow and Yu (2001) and was then extended to multivariate setting by Yu (2007). Under this model, the default likelihood of a firm is directly affected by the default status of other firms. See Appendix A.13 for a summary of the counterparty risk model.

In a related empirical study, Chava and Jarrow (2004) examined if industry effect plays a role in predicting the probability of a firm's bankruptcy. They divided the firms into four industrial groups according to SIC codes, and ran a logistic regression on each group of firms. Apart from a better in-sample fit, introducing the industrial factor significantly improves their out-of-sample forecast of bankruptcy events.

A robust line of research uses panel data techniques to study the default risk of firms. The default probabilities of firms are modeled by individual conditional intensity functions. A common way to model dependence of defaults among firms is to include exogenous factors that enter the default intensities of all firms. This type of *conditional independence models*, also known as *Cox models* or *doubly stochastic models*, is straightforward to estimate because the defaults of firms are independent of each other after controlling for exogenous factors. In

day span from July 5 to August 29, 2000.

a log-linear regression, Das, Duffie, Kapadia and Saita (2006, DDKS hereafter) estimate the default probabilities of U.S. firms over a 25 year time span (January 1979 to October 2004) with exogenous factors¹⁵. However, a series of diagnostic checks unanimously rejects the estimated DDKS model. A potential reason is an incorrect conditional independence assumption, but it could also be due to missing covariates. Their work stimulated future research effort in the pursuit of a more adequate default risk model. As a follow-up, Duffie, Eckners, Horel and Saita (2009) attempt to extend the DDKS model by including additional latent variables. Lando and Nielsen (2010) validate the conditional independence assumption by identifying another exogenous variable (industrial productivity index) and showing that the DDKS model with this additional covariate cannot be rejected.

In view of the inadequacy of conditional independence models, Azizpour, Giesecke and Schwenkler (2008) advocate a top-down approach to modeling corporate bankruptcies: rather than focusing on firm-specific default intensities, they directly model the aggregate default intensity for all firms over time. This approach offers a macroscopic view of default pattern of a portfolio of 6,048 issuers of corporate debts in the U.S.. A key advantage of this approach is that it provides a parsimonious way to model *self-exciting dynamics* which is hard to incorporate in the DDKS model. The authors showed that the self-exciting mechanism effectively explains a larger portion of default clustering. Idiosyncratic components such as firm-specific variables may indirectly drive the dynamics of the default process through the self-exciting mechanism.

¹⁵They include macroeconomic variables such as three-year Treasury yields and trailing one year return of S&P500 index, and firm-specific variables such as distance to default and trailing one year stock return.

2.1.5 Test of Dependence between two stochastic processes

Various techniques that test for the dependence between two stochastic processes are available. They are particularly well studied when the processes are time series in discrete time. Inspired by the seminal work of Box and Pierce (1970), Haugh (1976) derives the asymptotic distribution of the residual cross-correlations between two independent covariance-stationary ARMA models. A chi-squared test of no cross-correlation up to a fixed lag is constructed in the form of a sum of squared cross-correlations over a finite number of lags. Hong (1996b) generalizes Haugh's test by considering a weighted sum of squared cross-correlations over all possible lags, thereby ensuring consistency against all linear alternatives with significant cross-correlation at any lag. A similar test of serial dependence is developed for dynamic regression models with unknown forms of serial correlations (Hong, 1996a).

In the point process literature, there exist similar tests of no cross-correlation. Cox (1965) proposes an estimator of the second-order intensity function of a univariate stationary point process and derived the first two moments of the estimator when the process is a Poisson process. Cox and Lewis (1972) extend the estimator to a bivariate stationary point process framework. Brillinger (1976) derives the pointwise asymptotic distribution of the second-order intensity function estimator when the bivariate process exhibits no cross-correlation and satisfies certain mixing conditions. Based on these theoretical results, one can construct a test statistic in the form of a (weighted) summation of the second-order intensity estimator over a countable number of lags. Under the null of no cross-correlations, the test statistic has an asymptotic standard normal distribution. Doss (1991) considers the same testing problem but proposes

using the distribution function analog to the second-order intensity function as a test statistic. Under a different set of moment and mixing conditions, he shows that this test is more efficient than Brillinger's test while retaining asymptotic normality. Similar to the work of Brillinger, Doss' asymptotic normality result holds in a pointwise sense only. The users of these tests are left with the task of determining the grid of lags to evaluate the intensity function estimator. The grid of lags must be sparse enough to ensure independence so that central limit theorem is applicable, but not too sparse as to leave out too many alternatives. For the test considered in this chapter, such concern is removed because the test statistic is in the form of a weighted integration over a continuum of lags up to the sample length. The concept of Granger causality for point process is also related to independent censoring in the survival analysis literature.¹⁶

2.2 Bivariate Point Process

The bivariate point process $\mathbf{\Pi}$ consists of two sequences of event time $0 < \tau_1^k \leq \tau_2^k \leq \dots < \infty$ ($k = a, b$) on the positive real line \mathbb{R}_+ , where τ_i^k represents the time at which the i^{th} event of type k occurs. Another representation of the event time sequences is the bivariate counting process $\mathbf{N} = (N^a, N^b)'$, with the marginal counting process for type k events defined by $N^k(B) = \sum_{i=1}^{\infty} 1\{\tau_i^k \in B\}$, $k = a, b$, for any set B on \mathbb{R}_+ . Let $N_t^k = N^k((0, t])$ for all $t > 0$ and $N_0^k = 0$, $k = a, b$. It is clear that both representations are equivalent - from a trajectory of \mathbf{N} one can recover

¹⁶Independent censoring occurs when censoring times are independent of event times. If the bivariate point process is composed of censoring times and event times which are both observable, then we may check for pairwise independent censoring by means of the nonparametric Granger causality test to be studied in Section 2.4. A rejection of Granger non-causality between event and censoring times would then imply a rejection of pairwise independent censoring (as independence is a stronger assumption than Granger non-causality). Thanks to Robert Strawderman for pointing this out.

that of Π and vice versa; hence, for notational simplicity, the probability space for both Π and \mathbf{N} is denoted by (Ω, P) .

First, I suppose that the bivariate counting process \mathbf{N} satisfies the following assumption:

Assumption (A1) The *pooled* counting process $N \equiv N^a + N^b$ is simple, that is

$$P(N(\{t\}) = 0 \text{ or } 1 \text{ for all } t) = 1.$$

Essentially, assumption (A1) means that, almost surely, there is at most one event happening at any time point, and if an event happens, it can either be a type a or type b event, but not both. In other words, the pooled counting process N , which counts the number of events over time regardless of event types, is a monotonic increasing piecewise constant random function which jumps by exactly one at countable number of time points or otherwise stays constant at integer values. As it turns out, this simple property imposed on the pooled counting process plays a crucial role in simplifying the computation of moments of the test statistic. More importantly, the Bernoulli nature of the increments dN_t (which is either zero or one almost surely) of N at time t implies that if two increments dN_s and dN_t ($s \neq t$) are uncorrelated, then they must be *independent*.¹⁷ Therefore, a statistical test that checks for zero cross correlation between any pair of increments of N^a and N^b is sufficient for testing for pairwise independence between the increments.

In theory, assumption (A1) is mild enough to include a wide range of bivariate point process models. It is certainly satisfied if events happen randomly and

¹⁷If two random variables X and Y are uncorrelated, it does not follow in general that they are statistically independent. However, there are two exceptions: one is when (X, Y) follows a bivariate normal distribution, another is when X and Y are Bernoulli distributed.

independently of each other over a continuum (i.e. when the pooled point process is a Poisson process). Also, the assumption is often imposed on the pooled process of many other bivariate point process models that are capable of generating dependent events (e.g. doubly stochastic models, bivariate Hawkes models, bivariate autoregressive conditional intensity models). In practice, however, it is not uncommon to have events happening at exactly the same time point. In many cases, this is the artifact of recording or collecting point process data over a discrete time grid that is too coarse.¹⁸ In some other cases, multiple events really happen at the same time. Given a fixed time resolution, it is impossible to tell the difference between the two cases.¹⁹ There are two ways to get around this conundrum: I may either drop assumption (A1) and include a bigger family of models (e.g. compound Poisson processes), or keep the assumption but lump multiple events at the same time point into a single event. In this chapter, I would adopt the latter approach by keeping assumption (A1) and treating multiple events at the same time point as a single event, so that an occurrence of a type k event is interpreted as an occurrence of at least one type k event at that time point. In the datasets of empirical applications, the proportions that events of different types occur simultaneously turn out to be small or even zero by construction.²⁰

¹⁸For instance, in a typical TAQ dataset, timestamps for trades and quote revisions are accurate up to a second. There is a considerable chance that more than two transactions or quote revisions happen within a second. This is at odds with assumption (A1).

¹⁹TAQ datasets recorded with millisecond timestamps are available more recently. The improvement in resolution of timestamps mitigates the conflict with assumption (A1) by a large extent. A comparison with the TAQ datasets with timestamps in seconds can reveal whether a lump of events in the latter datasets is indeed the case or due to discrete time recording.

²⁰Among all trades and quote revisions of PG (GM) from 1997/8/4 to 1997/9/30 in the TAQ data, 3.6% (2.6%) of them occur within the same second. In the bankruptcy data ranging from January 1980 to June 2010, the proportion of cases in which bankruptcies of a manufacturing related firm and a financial related firm occur on the same date is 4.9% (out of a total of 892 cases). In the international financial contagion data, the proportions are all 0% because I intentionally pair up the leading indices of different stock markets which are in different time zones.

I can as well replace assumption (A1) by the assumption:

Assumption (A1b) the *pooled* counting process $N \equiv N^a + N^b$ is orderly, that is

$$P(N((0, s]) \geq 2) = o(s) \text{ as } s \downarrow 0.$$

It can be shown that with the second-order stationarity of N (see assumption (A2) to be stated later), assumptions (A1) and (A1b) are equivalent (Daley and Vere-Jones, 2003).

It is worth noting that assumptions (A1) and (A1b) are imposed on the pooled counting process N , and thus stronger than if they were imposed on the marginal processes N^a and N^b instead, because simple (or orderly) property of marginal counting processes does not carry over to the pooled counting process. For instance, if N^a is simple (or orderly) and $N^b \equiv N^a$ for each trajectory, then $N = N^a + N^b = 2N^a$ is not.

To make statistical inference possible, some sort of time homogeneity (i.e. stationarity) condition is necessary. Before discussing stationarity, let us define the *second-order factorial moment measure* as

$$G^{ij}(B_1 \times B_2) = E \left[\int_{B_2} \int_{B_1} 1_{\{t_1 \neq t_2\}} dN_{t_1}^i dN_{t_2}^j \right],$$

for $i, j = a, b$ (see Daley and Vere-Jones, 2003, section 8.1). Note that the indicator $1_{\{t_1 \neq t_2\}}$ is redundant if the pooled process of N is simple (assumption (A1)). The concept of second-order stationarity can then be expressed in terms of the second-order factorial moment measure $G^{ij}(\cdot, \cdot)$.

Definition 1 A bivariate counting process $\mathbf{N} = (N^a, N^b)'$ is second-order stationary if

$$(i) \ G^{ij}((0, 1]^2) = E \left[N^i((0, 1]) N^j((0, 1]) \right] < \infty \text{ for all } i, j = a, b; \text{ and}$$

(ii) $G^{ij}((B_1 + t) \times (B_2 + t)) = G^{ij}(B_1 \times B_2)$ for all bounded Borel sets B_1, B_2 in \mathbb{R}_+ and $t \in \mathbb{R}_+$.

The analogy to the stationarity concept in time series is clear from the above definition, which requires that the second-order (auto- and cross-) moments exist and that the second-order factorial moment measure is shift-invariant. By the shift-invariance property, the measure $G^{ij}(\cdot, \cdot)$ can be reduced to a function of one argument, say $\check{G}^{ij}(\cdot)$, as it depends only on the time difference of the component point process increments. If $\ell(\cdot)$ denotes the Lebesgue measure, then second-order stationarity of \mathbf{N} implies that, for any bounded measurable functions f with bounded support, the following decomposition is valid:

$$\int_{\mathbb{R}^2} f(s, t) G^{ij}(ds, dt) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, x + u) \ell(dx) \check{G}^{ij}(du).$$

From the moment condition in Definition 1 (i), second-order stationarity implies that the first-order moments exist by Cauchy-Schwarz inequality, so that

$$\lambda^k \equiv E[N^k((0, 1])] < \infty \quad (2.1)$$

for $k = a, b$. This is an integrability condition on N^k which ensures that events are not too closely packed together. Often known as *hazard rate* or *unconditional intensity*, the quantity λ^k gives the mean number of events from the component process N^k over a unit interval. Given stationarity, the unconditional intensity defined in (2.1) also satisfies $\lambda^k = \lim_{\Delta t \downarrow 0} (\Delta t)^{-1} P(N^k((t, t + \Delta t]) > 0)$. If I further assume that N^k is simple, then $\lambda = \lim_{\Delta t \downarrow 0} (\Delta t)^{-1} P(N^k((t, t + \Delta t]) = 1) = E(dN_t^k/dt)$, which is the mean occurrence rate of events at any time instant t , thus justifying the name *intensity*.

Furthermore, if the reduced measure $\check{G}^{ij}(\cdot)$ is absolutely continuous, then the reduced form factorial product densities $\varphi^{ij}(\cdot)$ ($i, j = a, b$) exist, so that, in

differential form, $\check{G}^{ij}(d\ell) = \varphi^{ij}(\ell) d\ell$. It is important to note that the factorial product density function $\varphi^{ij}(\ell)$ is not symmetric about zero unless $i = j$. Also, the reduced form auto-covariance (when $i = j$) and cross-covariance (when $i \neq j$) density functions of \mathbf{N} are well-defined:

$$c^{ij}(\ell) \equiv \varphi^{ij}(\ell) - \lambda^i \lambda^j \quad (2.2)$$

for $i, j = a, b$.

The assumptions are summarized as follows:

Assumption (A2) The bivariate counting process $\mathbf{N} = (N^a, N^b)$ is *second-order stationary* and that the *second-order reduced product densities* $\varphi^{ij}(\cdot)$ ($i, j = a, b$) exist.

Analogous to time series modeling, there is a strict stationarity concept: a bivariate process $\mathbf{N} = (N^a, N^b)$ is *strictly stationary* if the joint distribution of $\{\mathbf{N}(B_1 + u), \dots, \mathbf{N}(B_r + u)\}$ does not depend on u , for all bounded Borel sets B_i on \mathbb{R}^2 , $u \in \mathbb{R}^2$ and integers $r \geq 1$. Provided that the second-order moments exist, strict stationarity is stronger than second-order stationarity.

While the simple property is imposed on the pooled point process in assumption (A1), second-order stationarity is required for the bivariate process in assumption (A2). Suppose instead that only the pooled counting process is assumed second-order stationary. It does not follow that the marginal counting processes are second-order stationary too.²¹

²¹For instance, if $N = N^a + N^b$ is second-order stationary, and if we define $N_t^a = N(\cup_{i \geq 0} (2i, 2i+1] \cap (0, t])$ and $N_t^b = N_t - N_t^a$, then N^a and N^b are clearly not second-order stationary. The statement is still valid if second-order stationarity is replaced by strict stationarity.

The assumption of second-order stationarity on \mathbf{N} ensures that the mean and variance of the test statistic (to be introduced in (2.13)) are finite under the null hypothesis of no causality (in (2.11)). Nevertheless, it is possible to accommodate non-stationary point processes with time-varying unconditional intensities by relaxing assumption (A2). See Appendix A.14 for the asymptotic theory of the test for this case.

In order to show asymptotic normality I need to assume the existence of fourth-order moments for each component process, as follows:

Assumption (A6) $E \left[\{N^k(B_1)N^k(B_2)N^k(B_3)N^k(B_4)\} \right] < \infty$ for $k = a, b$ and for all bounded Borel sets B_i on \mathbb{R}_+ , $i = 1, 2, 3, 4$.

Fourth-order moment condition is typical for invoking central limit theorems. In a related work, David (2008) imposes a much stronger assumption of Brillinger-mixing, which essentially requires the existence of all moments of the point process over bounded intervals.

Before proceeding, let me introduce another important concept: the *conditional intensity* of a counting process:

Definition 2 Given a filtration²² $\mathcal{G} = (\mathcal{G}_t)_{t \geq 0}$, the \mathcal{G} -conditional intensity $\lambda(t|\mathcal{G}_{t-})$ of a univariate counting process $\check{N} = (\check{N}_t)_{t \geq 0}$ is any \mathcal{G} -measurable stochastic process such that for any Borel set B and any \mathcal{G}_t -measurable function C_t , the following condition is satisfied:

$$E \left[\int_B C_t d\check{N}_t \right] = E \left[\int_B C_t \lambda(t|\mathcal{G}_{t-}) dt \right]. \quad (2.3)$$

²²All filtrations in this chapter satisfy the usual conditions in Protter (2004).

It can be shown (Brémaud, 1981) that the \mathcal{G} -conditional intensity $\lambda(t|\mathcal{G}_{t-})$ is unique almost surely if those $\lambda(t|\mathcal{G}_{t-})$ that satisfy (2.3) are required to be \mathcal{G} -predictable. In the rest of the chapter, I will assume predictability for all conditional intensity functions (see assumption (A3) at the end of this section).

Similar to unconditional intensity, we can interpret the conditional intensity at time t of a simple counting process \check{N} as the mean occurrence rate of events given the history \mathcal{G} just before time t , as $\lambda(t|\mathcal{G}_{t-}) = \lim_{\Delta t \downarrow 0} (\Delta t)^{-1} P(\check{N}((t, t + \Delta t]) > 0 | \mathcal{G}_{t-}) = \lim_{\Delta t \downarrow 0} (\Delta t)^{-1} P(\check{N}((t, t + \Delta t]) = 1 | \mathcal{G}_{t-}) = E(d\check{N}_t/dt | \mathcal{G}_{t-})$, P -almost surely²³, where the second equality follows from (A1).

Let $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ be the natural filtration of the bivariate counting process \mathbf{N} , i.e. $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$, and $\mathcal{F}^k = (\mathcal{F}_t^k)_{t \geq 0}$ ($k = a, b$) be the natural filtration of N^k , so that \mathcal{F}_t and \mathcal{F}_t^k are the sigma fields generated by the processes \mathbf{N} and N^k on $[0, t]$, i.e. $\mathcal{F}_t = \sigma\{(N_s^a, N_s^b), 0 \leq s \leq t\}$ and $\mathcal{F}_t^k = \sigma\{N_s^k : s \in [0, t]\}$. Clearly, $\mathcal{F} = \mathcal{F}^a \vee \mathcal{F}^b$. Let $\lambda^k(t|\mathcal{F}_{t-})$ be the \mathcal{F} -conditional intensity of N_t^k , and define the error process by

$$e_t^k := N_t^k - \int_0^t \lambda^k(s|\mathcal{F}_{s-}) ds \quad (2.4)$$

for $k = a, b$.

By Doob-Meyer decomposition, the error process e_t^k is an \mathcal{F} -martingale process, in the sense that $E(e_t^k | \mathcal{F}_s) = e_s^k$ for all $t > s \geq 0$. The integral $\Lambda_t^k = \int_0^t \lambda^k(s|\mathcal{F}_{s-}) ds$ as a process is called the \mathcal{F} -compensator of N_t^k which always exists by Doob-Meyer decomposition, but the existence of \mathcal{F} -conditional intensity $\lambda^k(t|\mathcal{F}_{t-})$ is not guaranteed unless the compensator is *absolutely continuous*. For later analyses, I will assume the existence of $\lambda^k(t|\mathcal{F}_{t-})$ (see assumption (A3) at the end of this section).

²³In the rest of the chapter, all equalities involving conditional expectations hold in an almost surely sense.

I can express (2.4) in differential form:

$$de_t^k = dN_t^k - \lambda^k(t|\mathcal{F}_{t-})dt = dN_t^k - E(dN_t^k|\mathcal{F}_{t-})$$

for $k = a, b$. From the martingale property of e_t^k , it is then clear that the differential de_t^k is a mean-zero martingale process. In particular, $E(de_t^k|\mathcal{F}_{t-}) = 0$ for all $t > 0$. In other words, based on the bivariate process history \mathcal{F}_{t-} just before time t , an econometrician can obtain the \mathcal{F} -conditional intensities $\lambda^a(t|\mathcal{F}_{t-})$ and $\lambda^b(t|\mathcal{F}_{t-})$ which are computable just before time t (recall that $\lambda^k(t|\mathcal{F}_{t-})$ is \mathcal{F} -predictable) and give the best prediction of the bivariate counting process \mathbf{N} at time t . Since by (A1) the term $\lambda^k(t|\mathcal{F}_{t-})dt$ becomes the conditional mean of dN_t^k , the prediction is the best in the mean square sense.

One may wonder whether it is possible to achieve equally accurate prediction of $\mathbf{\Pi}$ with a reduced information set. For instance, can we predict dN_t^b equally well with its \mathcal{F}^b -conditional intensity $\lambda^b(t|\mathcal{F}_{t-}^b)$, where $\lambda^b(t|\mathcal{F}_{t-}^b)dt = E(dN_t^b|\mathcal{F}_{t-}^b)$, instead of its \mathcal{F} -conditional intensity $\lambda^b(t|\mathcal{F}_{t-})$? Through computing the \mathcal{F}^b -conditional intensity, we attempt to predict the value of N^b solely based on the history of N^b . Without using the history of N^a , the prediction $\lambda^b(t|\mathcal{F}_{t-}^b)dt$ ignores the feedback or causal effect that shocks to N^a in the past may have on the future dynamics of N^b . One would thus expect the answer to the previous question is no in general. Indeed, given that $\mathbf{\Pi}$ is in the filtered probability space (Ω, P, \mathcal{F}) , the error process

$$\epsilon_t^b := N_t^b - \int_0^t \lambda^b(s|\mathcal{F}_{s-}^b)ds \quad (2.5)$$

is no longer an \mathcal{F} -martingale. However, ϵ_t^b is an \mathcal{F} -martingale under one special circumstance: when the \mathcal{F}^b - and \mathcal{F} -conditional intensities

$$\lambda^b(t|\mathcal{F}_{t-}^b) = \lambda^b(t|\mathcal{F}_{t-})$$

are the same for all $t > 0$. I am going to discuss this circumstance in depth in the next section.

Let me summarize the assumptions in this section:

Assumption (A3) The \mathcal{F} -conditional intensity $\lambda^k(t|\mathcal{F}_{t-})$ and \mathcal{F}^k -conditional intensity $\lambda_t^k \equiv \lambda^k(t|\mathcal{F}_{t-}^k)$ of the counting process N_t^k exist and are predictable.

2.3 Granger Causality

In this section, I am going to discuss the concept of Granger causality in the bivariate counting process set-up described in the previous section. Assume (A1), (A2) and (A3) and the notations in the previous section. Then, for two distinct marginal counting processes N^a and N^b , we say that N^a *does not Granger-cause* N^b if the \mathcal{F} -conditional intensity of N^b is identical to the \mathcal{F}^b -conditional intensity of N^b . That is, for all $t > s \geq 0$, P -almost surely,

$$E[dN_t^b|\mathcal{F}_s] = E[dN_t^b|\mathcal{F}_s^b] \quad (2.6)$$

A remarkable result, as proven by Florens and Fougere (1996, section 4, example I), is the following equivalence statement in the context of simple counting processes.

Theorem 3 *If N^a and N^b are simple counting processes, then the following four definitions of Granger noncausality are equivalent:*

1. N^a does not weakly globally cause N^b , i.e. $E[dN_t^b|\mathcal{F}_s] = E[dN_t^b|\mathcal{F}_s^b]$, P -a.s. for all s, t .

2. N^a does not strongly globally cause N^b , i.e. $\mathcal{F}_t^b \perp \mathcal{F}_s | \mathcal{F}_s^b$ for all s, t .
3. N^a does not weakly instantaneously cause N^b , i.e. N^b , which is an \mathcal{F}^b -semi-martingale with decomposition $dN_t^b = d\epsilon_t^b + E[dN_t^b | \mathcal{F}_t^b]$, remains an \mathcal{F} -semi-martingale with the same decomposition.
4. N^a does not strongly instantaneously cause N^b , i.e. any \mathcal{F}^b -semi-martingale with decomposition remains an \mathcal{F} -semi-martingale with the same decomposition.

According to the theorem, weakly global noncausality is equivalent to weakly instantaneous noncausality, and hence testing for (2.6) is equivalent to checking ϵ_t^b defined in (2.5) is an \mathcal{F} -martingale process, or, checking $d\epsilon_t^b$ is an \mathcal{F} -martingale difference process:

$$E[d\epsilon_t^b | \mathcal{F}_s] = 0 \quad (2.7)$$

for all $0 \leq s < t$.

If one is interested in testing for pairwise dependence only, then (2.7) implies

$$E[f(d\epsilon_s^a) d\epsilon_t^b] = 0 \quad (2.8)$$

and

$$E[f(d\epsilon_s^b) d\epsilon_t^b] = 0 \quad (2.9)$$

for all $0 \leq s < t$ and any \mathcal{F}^a -measurable function $f(\cdot)$. However, since ϵ_t^b is an \mathcal{F}^b -martingale by construction, condition (2.9) is automatically satisfied and thus is not interesting from testing's point of view as long as the conditional intensity $\lambda^b(t | \mathcal{F}_t^b)$ is computed correctly.

There is a loss of generality to base a statistical test on (2.8) instead of (2.7), as it would miss the alternatives in which a type b event is not Granger-caused by

the occurrence (or non-occurrence) of any single type a event at a past instant, but is Granger-caused by the occurrence (or non-occurrence) of multiple type a events jointly at multiple past instants or over some past intervals.²⁴

I can simplify the test condition (2.8) further. Due to the dichotomous nature of $d\epsilon_t^a$, it suffices to test

$$E[d\epsilon_s^a d\epsilon_t^b] = 0 \quad (2.10)$$

for all $0 \leq s < t$, as justified by the following lemma.

Lemma 4 *If N^a and N^b are simple counting processes, then (2.8) and (2.10) are equivalent.*

Proof. The implication from (2.8) to (2.10) is trivial by taking $f(\cdot)$ to be the identity function. Now assuming that (2.10) holds, i.e. $\text{Cov}(d\epsilon_s^a, d\epsilon_t^b) = 0$. Given that N^a and N^b are simple, $dN_s^a|\mathcal{F}_{s^-}^a$ and $dN_t^b|\mathcal{F}_{t^-}^b$ are Bernoulli random variables (with means $\lambda^a(s|\mathcal{F}_{s^-}^a)ds$ and $\lambda^b(t|\mathcal{F}_{t^-}^b)dt$, respectively), and hence zero correlation implies independence, i.e. for all measurable functions $f(\cdot)$ and $g(\cdot)$, we have $\text{Cov}[f(d\epsilon_s^a), g(d\epsilon_t^b)] = 0$. We thus obtain (2.8) by taking $g(\cdot)$ to be the identity function. ■

Thanks to the simple property of point process assumed in (A1), two innovations $d\epsilon_s^a$ and $d\epsilon_t^b$ are pairwise cross-independent if they are not pairwise cross-correlated by Lemma 4. In other words, a suitable linear measure of cross-correlation between the residuals from two component processes would suffice

²⁴One hypothetical example in default risk application is given as follows. Suppose I want to detect whether corporate bankruptcies in industry a Granger-cause bankruptcies in industry b . Suppose also that there were three consecutive bankruptcies in industry a at times s_1, s_2 and s_3 , followed by a bankruptcy in industry b at time t ($s_1 < s_2 < s_3 < t$). Each bankruptcy in industry a alone would not be significant enough to influence the well-being of the companies in industry b , but three industry a bankruptcies may jointly trigger an industry b bankruptcy. It is possible that a test based on (2.8) can still pick up such a scenario, depending on the way the statistic summarizes the information of (2.8) for all $0 \leq s_i < t$.

to test for their pairwise cross-independence (both linear and nonlinear), as each infinitesimal increment takes one out of two values almost surely. From testing's point of view, a continuous time framework justifies the simple property of point processes (assumption (A1)) and hence allows for a simpler treatment on the nonlinearity issue, as assumption (A1) gets rid of the possibility of nonlinear dependence on the infinitesimal level. Indeed, if a point process \check{N} is simple, then $d\check{N}_t$ can only take values zero (no jump at time t) or one (a jump at time t), and so $(d\check{N}_t)^p = d\check{N}_t$ for any positive integers p . Without assumption (A1), the test procedure would still be valid (to be introduced in section 2.4, with appropriate adjustments to the mean and variance of the test statistic), but it would just check for an implication of pairwise Granger noncausality, as the equivalence of (2.8) and (2.10) would be lost.

Making sense of condition (2.10) requires a thorough understanding of the conditional intensity concept and its relation to Granger causality. From Definition 2, it is crucial to specify the filtration with respect to which the conditional intensity is adapted. The \mathcal{G} -conditional intensity can be different depending on the choice of the filtration \mathcal{G} . If $\mathcal{G} = \mathcal{F} = \mathcal{F}^a \vee \mathcal{F}^b$, then the \mathcal{G} -conditional intensity is evaluated with respect to the history of the whole bivariate counting process \mathbf{N} . If instead $\mathcal{G} = \mathcal{F}^k$, then it is evaluated with respect to the history of the marginal point process N^k only.

From the definition of weakly instantaneous noncausality in Theorem 3, Granger-noncausality for point processes is the property that the conditional intensity is invariant to an enlargement of the conditioning set from the natural filtration of the marginal process to that of the bivariate process. More specifi-

cally, if the counting process N^a does not Granger-cause N^b , then we have

$$E[dN_t^b | \mathcal{F}_{t-}] = E[dN_t^b | \mathcal{F}_{t-}^b]$$

for all $t > 0$, which conforms to the intuition of Granger causality that the predicted value of N_t^b given its history remains unchanged with or without the additional information of the history of N^a by time t . Condition (2.10), on the other hand, means that any past innovation $d\epsilon_s^a = dN_s^a - E[dN_s^a | \mathcal{F}_{s-}^a]$ of N^a is *independent of* (not merely uncorrelated with, due to the Bernoulli nature of jump sizes for simple point processes according to Lemma 4) the future innovation $d\epsilon_t^b = dN_t^b - E[dN_t^b | \mathcal{F}_{t-}^b]$ of N^b ($t > s$). This is exactly the implication of Granger noncausality from N^a to N^b , and except for those loss-of-generality cases discussed underneath (2.9), the two statements are equivalent.

Assuming (A2) and (A3), the *reduced form cross covariance density function* of the innovations $d\epsilon_t^a$ and $d\epsilon_t^b$ is then well-defined, and is denoted by $\gamma(\ell) dt d\ell = E(d\epsilon_t^a d\epsilon_{t+\ell}^b)$. The null hypothesis of interest can thus be written down formally as follows:

$$\mathbf{H}_0 : \gamma(\ell) = 0 \text{ for all } \ell > 0 \quad \text{vs} \quad (2.11)$$

$$\mathbf{H}_1 : \gamma(\ell) \neq 0 \text{ for some } \ell > 0.$$

It is important to distinguish the reduced form cross-covariance density function $\gamma(\ell)$ of the innovations $d\epsilon_t^a$ and $d\epsilon_t^b$ from the cross-covariance density function $c^{ab}(\ell)$ of the counting process $\mathbf{N} = (N^a, N^b)$, defined earlier in (2.2). The key difference rests on the way the jumps are demeaned: the increment dN_t^k at time t is compared against the conditional mean $\lambda^k(t | \mathcal{F}_{t-}^k) dt$ in $\gamma(\ell)$, but it is compared against the unconditional mean $\lambda^k dt$ in $c^{ab}(\ell)$. In this sense, the former $\gamma(\ell)$ captures the *dynamic* feedback effect as reflected in the shocks of the

component processes, but the latter $c^{ab}(\ell)$ merely summarizes the *static* correlation relationship between the jumps of component processes. Indeed, valuable information of Granger causality between component processes is only contained in $\gamma(\ell)$ (as argued earlier in this section) but not in $c^{ab}(\ell)$. Previous research focused mostly on the large sample properties of estimators of the static auto-covariance density function $c^{kk}(\ell)$ or cross-covariance density function $c^{ab}(\ell)$. This chapter, however, is devoted to the analysis of the dynamic cross-covariance density function $\gamma(\ell)$. As we will see, the approach in getting asymptotic properties of $\gamma(\ell)$ is quite different. I will apply the *martingale central limit theorem* - a dynamic version of the ordinary central limit theorem - to derive the sampling distribution of a test statistic involving estimators of $\gamma(\ell)$.

2.4 The statistic

The econometrician observes two event time sequences of a simple bivariate stationary point process $\mathbf{\Pi}$ over the time horizon $[0, T]$, namely, $0 < \tau_1^k < \tau_2^k < \dots < \tau_{N^k(T)}^k$ for $k = a, b$. This is the dataset required to calculate the test statistic to be constructed in this section.

2.4.1 Nonparametric cross-covariance estimator

In this section, I am going to construct a statistic for testing condition (2.10) from the data. One candidate for the lag ℓ sample cross-covariance $\gamma(\ell)$ of the innovations $d\epsilon_t^a$ and $d\epsilon_t^b$ is given by

$$\hat{C}(\ell)d\ell = \frac{1}{T} \int_0^T d\hat{\epsilon}_t^a d\hat{\epsilon}_{t+\ell}^b$$

where $d\hat{\epsilon}_t^k = dN_t^k - \hat{\lambda}_t^k dt$ ($k = a, b$) is the residual and $\hat{\lambda}_t^k$ is some local estimator of the \mathcal{F}^k -conditional intensity λ_t^k in (A3) (to be discussed in section 2.4.4). The integration is done with respect to t . However, if the jumps of N^k are finite or countable (which is the case for point processes satisfying (A2)), the product of increments $dN_t^a dN_{t+\ell}^b$ is zero almost everywhere except over a set of P -measure zero, so that $\hat{C}(\ell)$ is inconsistent for $\gamma(\ell)$. This suggests that some form of local smoothing is necessary. The problem is analogous to the probability density function estimation in which the empirical density estimator would be zero almost everywhere over the support if there were no smoothing. This motivates the use of a kernel function $K(\cdot)$, with a bandwidth H which controls the degree of smoothing applied to the sample cross-covariance estimator $\hat{C}(\ell)$ above. To simplify notation, let $K_H(x) = K(x/H)/H$. The corresponding kernel estimator is given by

$$\begin{aligned}\hat{\gamma}_H(\ell) &= \frac{1}{T} \int_0^T \int_0^T K_H(t-s-\ell) d\hat{\epsilon}_s^a d\hat{\epsilon}_t^b \\ &= \frac{1}{T} \int_0^T \int_0^T K_H(t-s-\ell) (dN_s^a - \hat{\lambda}_s^a ds) (dN_t^b - \hat{\lambda}_t^b dt).\end{aligned}\tag{2.12}$$

The kernel estimator is the result of averaging the weighted products of innovations $d\hat{\epsilon}_s^a$ and $d\hat{\epsilon}_t^b$ over all possible pairs of time points (s, t) . The kernel $K_H(\cdot)$ gives the heaviest weight to the product of innovations at the time difference $t - s = \ell$, and the weight becomes lighter as the time difference is further away from ℓ . The following integrability conditions are imposed on the kernel:

Assumption (A4a) The kernel function $K(\cdot)$ is symmetric around zero and satisfies $\kappa_1 \equiv \int_{-\infty}^{\infty} K(u) du = 1$, $\kappa_2 \equiv \int_{-\infty}^{\infty} K^2(u) du < \infty$, $\kappa_4 \equiv \iiint_{(-\infty, \infty)} K(u) K(v) K(u + w) K(v + w) du dv dw < \infty$ and $\int_{-\infty}^{\infty} u^2 K(u) du < \infty$.

2.4.2 The statistic as \mathcal{L}^2 norm

An ideal test statistic for testing (2.11) would summarize appropriately all the cross-covariances of residuals $d\hat{\epsilon}_s^a$ and $d\hat{\epsilon}_t^b$ over all $0 \leq s < t$. This problem is similar to that of Haugh (1976) when he checked the independence of two time series, but there are two important departures: here I am working with two continuous time point processes instead of discrete time series, and I do not assume any parametric models on the conditional means. To this end, I propose a weighted integral of the squared sample cross-covariance function, defined as follows:

$$Q \equiv \|\hat{\gamma}_H\|_2 \equiv \int_I w(\ell) \hat{\gamma}_H^2(\ell) d\ell. \quad (2.13)$$

where $I \subseteq [-T, T]$. To test the null hypothesis in (2.11), the integration range is set to be $I = [0, T]$.

Applying an \mathcal{L}^2 norm rather than an \mathcal{L}^1 norm on the sample cross-covariance function $\hat{\gamma}_H(\ell)$ is standard in the literature of discrete time serial correlation test. If I decided to test (2.11) based on

$$\|\hat{\gamma}_H\|_1 \equiv \int_I w(\ell) \hat{\gamma}_H(\ell) d\ell$$

instead, it would lead to excessive type II error - the test would fail to reject those DGP's in which the true cross-covariance function $\gamma(\ell)$ is significantly away from zero for certain $\ell \in I$ but the weighted integral $\|\hat{\gamma}_H\|_1$ is close to zero due to cancellation.

A test based on the test statistic Q in (2.13) is on the conservative side as Q is an \mathcal{L}^2 norm.²⁵ More specifically, the total causality effect from N^a to N^b is

²⁵In addition, the Q test has zero power against alternatives in which $\gamma(\ell) \neq 0$ for ℓ in some sets of measure zero.

the aggregate of the weighted squared contribution from each individual type a -type b event pair (see Figure A.2). However, less conservative test can be constructed with other choices of norms (e.g. Hellinger and Kullback-Leibler distance) as in Hong (1996a), and the methodology in this chapter is still valid with appropriate adjustment.

2.4.3 Weighting function

I assume that

Assumption (A5) The weighting function $w(\ell)$ is integrable over $(-\infty, \infty)$:

$$\int_{-\infty}^{\infty} w(\ell) d\ell < \infty.$$

The motivations behind the introduction of the weighting function $w(\ell)$ on lags are in a similar spirit as the test of serial correlation proposed by Hong (1996a) in the discrete time series context. The economic motivation is that the contagious effect from one process to another diminishes over time, as manifested by the property that the weighting function discounts more heavily the sample cross covariance as the time lag increases. From the econometric point of view, by choosing a weighting function whose support covers all possible lags in $I \subseteq [-T, T]$, the statistic Q can deliver a consistent test to (2.11) against all pairwise cross dependence of the two processes as it summarizes their cross covariances over all lags in an \mathcal{L}^2 norm, whereas the statistic with a truncated weighting function over a fixed lag window $I = [c_1, c_2]$ cannot. From the statistical point of view, a weighting function that satisfies (A5) is a crucial device for controlling the variation of the integrated squared cross-covariance function

over an expanding lag interval $I = [0, T]$, so that Q enjoys asymptotic normality. It can be shown that the asymptotic normality property would break down without an appropriate weighting function $w(\ell)$ that satisfies (A5).

2.4.4 Conditional intensity estimator

In this section, I will discuss how to estimate the time-varying \mathcal{F}^k -conditional intensity nonparametrically. I employ the following *Nadaraya-Watson estimator* for the \mathcal{F}^k -conditional intensity $\lambda_t^k \equiv \lambda^k(t|\mathcal{F}_t^k)$,

$$\hat{\lambda}_t^k = \int_0^T \hat{K}_M(t-u) dN_u^k. \quad (2.14)$$

While the cross-covariance estimator $\hat{\gamma}_H(\ell)$ is smoothed by the kernel $K(\cdot)$ with bandwidth H , the conditional intensity estimator is smoothed by the kernel $\hat{K}(\cdot)$ with bandwidth M . The kernel $\hat{K}(\cdot)$ is assumed to satisfy the following:

Assumption (A4b) The kernel function $\hat{K}(\cdot)$ is symmetric around zero and sat-

$$\text{isfies } \hat{\kappa}_1 \equiv \int_{-\infty}^{\infty} \hat{K}(u) du = 1, \hat{\kappa}_2 \equiv \int_{-\infty}^{\infty} \hat{K}^2(u) du < \infty, \hat{\kappa}_4 \equiv \iiint_{(-\infty, \infty)} \hat{K}(u) \hat{K}(v) \hat{K}(u+w) \hat{K}(v+w) dudvdw < \infty \text{ and } \int_{-\infty}^{\infty} u^2 \hat{K}(u) du < \infty.$$

The motivation of (2.14) comes from estimating the conditional mean of dN_t^k by a nonparametric local regression. Indeed, the Nadaraya-Watson estimator is the local constant least square estimator of $E(dN_t^k|\mathcal{F}_t^k)$ around time t weighted by $\hat{K}_M(\cdot)$. (As usual, I denote $\hat{K}_M(\ell) = \hat{K}(\ell/M)/M$.) By (A4b) it follows that $\int_0^T \hat{K}_M(t-u) du = 1 + o(1)$ as $M/T \rightarrow 0$ and thus the Nadaraya-Watson estimator becomes (2.14). The estimator (2.14) implies that the conditional intensity takes a constant value over a local window, but one may readily extend it to a local

linear or local polynomial estimator. Some candidates for regressors include the backward recurrence time $t - t_{N_t^k}^k$ of the marginal process N^k , and the backward recurrence time $t - t_{N_t}$ of the pooled process N .

Another way to estimate the \mathcal{F}^k -conditional intensity is by fitting a parametric conditional intensity model on each component point process. For $k = a, b$, let $\theta^k \in \mathbb{R}^{d^k}$ be the vector of parameters of the \mathcal{F}^k -conditional intensity λ_t^k , which is modeled by

$$\lambda_t^k \equiv \lambda^k(t; \theta^k)$$

for $t \in [0, \infty)$. Each component model is estimated by some parametric model estimation techniques (e.g. MLE, GMM). The estimator $\hat{\theta}^k$ converges to θ^k at the typical parametric convergence rate of $T^{-1/2}$ (or equivalently $(n^k)^{-1/2} = (N_T^k)^{-1/2}$), which is faster than the nonparametric rate of $M^{-1/2}$.

An advantage of parametric modeling is that *a-priori* dependence structure across the marginal point processes may be imposed. With the parametric specifications of the conditional intensities, the result of the nonparametric test based on Q in (2.13) would then detect the Granger causal feedback that is not explained by the parametric model. For instance, common seasonality $S(t)$ (deterministic) and covariates $X(t)$ (random and observable) may be introduced, as well as instantaneous latent factors $\eta(t)$ (random but unobservable) affecting both point processes, by setting $\lambda^k(t; \theta^k) = \exp\{S(t) + X(t) + \eta(t)\}$. This approach can be readily extended to a semi-parametric setting (in Cox's sense) by coupling a parametric conditional intensity model with a nonparametric estimator.

2.4.5 Computation of $\hat{\gamma}_H(\ell)$

To implement the test, it is important to compute the test statistic Q efficiently. From the definition, there are three layers of integrations to be computed: the first layer is the weighted integration with respect to different lags ℓ , a second layer involves two integrations with respect to the component point processes in the cross-covariance function estimator $\hat{\gamma}_H(\ell)$, and a third layer is a single integration with respect to each component process inside the \mathcal{F}^k -conditional intensity estimator $\hat{\lambda}_t^k$. The first layer of integration will be evaluated numerically, but it is possible to reduce the second and third layers of integrations to summations over marked event times in the case of Gaussian kernels, thus simplifying a lot the computation of $\hat{\gamma}_H(\ell)$ and hence Q . Therefore, I make the following assumption:

Assumption (A4d) The kernels $K(x)$, $\mathring{K}(x)$ and $\ddot{K}(x)$ are all standard Gaussian kernels. That is: $K(x) = \mathring{K}(x) = \ddot{K}(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.

Theorem 5 Under assumptions (A1-3, 4a, 4b, 4d), the cross-covariance function estimator $\hat{\gamma}_H(\ell)$ defined in (2.12) and (2.14) is given by

$$\hat{\gamma}_H(\ell) = \frac{1}{T} \sum_{i=1}^{N_T^a} \sum_{j=1}^{N_T^b} \left[\frac{1}{H} K\left(\frac{t_j^b - t_i^a - \ell}{H}\right) - \frac{2}{\sqrt{H^2 + M^2}} K\left(\frac{t_j^b - t_i^a - \ell}{\sqrt{H^2 + M^2}}\right) + \frac{1}{\sqrt{H^2 + 2M^2}} K\left(\frac{t_j^b - t_i^a - \ell}{\sqrt{H^2 + 2M^2}}\right) \right].$$

2.4.6 Consistency of conditional intensity estimator

Unlike traditional time series asymptotic theories in which data points are separated by a fixed (but possibly irregular) time lag in an expanding observation window $[0, T]$ (scheme 1), consistent estimation of moments of point processes

requires a fixed observation window $[0, T_0]$ in which events grow in number and are increasingly packed (scheme 2). The details of the two schemes are laid out in Table A.1.

As we will see shortly, the asymptotic mechanism of scheme 2 is crucial for consistent estimation of the first and second order moments, including the \mathcal{F}^k -conditional intensity functions λ_t^k for $k = a, b$, the auto- and cross-covariance density functions $c^{ij}(\cdot)$ of \mathbf{N} (for $i, j = a, b$), as well as the cross-covariance density function $\gamma(\cdot)$ of the innovation processes $d\epsilon_t^k$ for $k = a, b$. However, the limiting processes of scheme 2 would inadvertently distort various moments of \mathbf{N} . For instance, the \mathcal{F}^k -conditional intensity λ_t^k will diverge to infinity as the number of observed events $n^k = N^k(T_0)$ in a finite observation window $[0, T_0]$ goes to infinity. In contrast, under traditional time series asymptotics (scheme 1) as $T \rightarrow \infty$, the moment features of \mathbf{N} are maintained as the event times are fixed with respect to T , but all moment estimators are doomed to be pointwise inconsistent since new information is only added to the right of the process (rather than everywhere over the observation window) as $T \rightarrow \infty$.

Let us take the estimation of \mathcal{F}^k -conditional intensity function λ_t^k as an example. At first sight, scheme 1 is preferable because the spacing between events is fixed relative to the sample size and we want the conditional intensity λ_t^k at time t to be invariant to the sample size in the limit. However, the estimated \mathcal{F}^k -conditional intensity is not pointwise consistent under scheme 1's asymptotics since there are only a fixed and finite number of observations around time t . On the other hand, under scheme 2's asymptotics, the number of observations around any time t increases as the sample grows, thus ensuring consistent estimation of λ_t^k , but as events get more and more crowded in a local window

around time t , the \mathcal{F}^k -conditional intensity λ_t^k diverges to infinity.²⁶

How can we solve the above dilemma? Knowing that there is no hope to estimate λ_t^k consistently at each time t , let us stick to scheme 2, and estimate the moment properties of a rescaled counting process $\tilde{\mathbf{N}}_v = (\tilde{N}_v^a, \tilde{N}_v^b)$, where

$$\tilde{N}_v^k := \frac{N_{Tv}^k}{T} \quad (2.15)$$

for $k = a, b$ and $v \in [0, 1]$ (a fixed interval, with $T_0 = 1$). The stationarity property of $\tilde{\mathbf{N}}$ and the Bernoulli nature of the increments of the pooled process $\tilde{N} = \tilde{N}^a + \tilde{N}^b$ are preserved.²⁷ The time change acts as a bridge between the two schemes - the asymptotics of original process \mathbf{N} is governed by scheme 1, while that of the rescaled process $\tilde{\mathbf{N}}$ is governed by scheme 2; and the two schemes are equivalent to one another after rescaling by $1/T$. Indeed, it is easily seen, by a change of variable $t = Tv$, that the conditional intensities of \tilde{N}_v^k and N_{Tv}^k are identical:

$$\begin{aligned} \lambda_{Tv}^k &= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} E \left(N_{Tv+\Delta t}^k - N_{Tv}^k | \mathcal{F}_{Tv^-}^k \right) \\ &= \lim_{\Delta v \downarrow 0} \frac{1}{T \Delta v} E \left(N_{T(v+\Delta v)}^k - N_{Tv}^k | \mathcal{F}_{Tv^-}^k \right) \\ &= \lim_{\Delta v \downarrow 0} \frac{1}{\Delta v} E \left(\tilde{N}_{v+\Delta v}^k - \tilde{N}_v^k | \tilde{\mathcal{F}}_v^k \right) =: \tilde{\lambda}_v^k, \end{aligned} \quad (2.16)$$

where I denoted the natural filtration of \tilde{N}^k by $\tilde{\mathcal{F}}^k$ and the $\tilde{\mathcal{F}}^k$ -conditional intensity function of \tilde{N}_v^k by $\tilde{\lambda}_v^k$ on the last line.

If the conditional intensity $\tilde{\lambda}_v^k$ of the rescaled point process \tilde{N}_v^k is continuous and is an unknown but deterministic function, then it can be consistently estimated for each $v \in [0, 1]$. In the same vein, other second-order moments of $\tilde{\mathbf{N}}$ are well-defined and can be consistently estimated, including the (auto- and cross-)

²⁶Note the similarity of the problem to probability density function estimation on a bounded support.

²⁷Strictly speaking, the pooled process \tilde{N} of $\tilde{\mathbf{N}}$ is no longer simple because the increment $d\tilde{N}_t$ takes values of either zero or $1/T$ (instead of 1) almost surely, but the asymptotic theory of the test statistic on $\tilde{\mathbf{N}}$ only requires that the increments $d\tilde{N}_t^k$ are Bernoulli distributed with mean $\lambda_t^k dt$.

covariance density functions $\tilde{c}^{ij}(\cdot)$ of $\tilde{\mathbf{N}}$ (for $i, j = a, b$) and the cross-covariance density function $\tilde{\gamma}(\cdot)$ of the innovation processes $d\tilde{\epsilon}_v^k := d\tilde{N}_v^k - \tilde{\lambda}_v^k dv$ for $k = a, b$. Specifically, it can be shown that

$$\tilde{\gamma}(\sigma) = \gamma(T\sigma) \quad (2.17)$$

and $\tilde{c}^{ij}(\sigma) = c^{ij}(T\sigma)$ for $i, j = a, b$, and consistent estimation is possible for fixed $\sigma \in [0, 1]$.

To show the consistency and asymptotic normality of the conditional intensity kernel estimator $\hat{\lambda}_{T_v}^k$, the following assumption is imposed:

Assumption (A7) The rescaled counting process $\tilde{N}_u^k \equiv N_{Tu}^k/T$ (with natural filtration $\tilde{\mathcal{F}}^k$) has an $\tilde{\mathcal{F}}^k$ -conditional intensity function $\tilde{\lambda}_u^k$, which is twice continuously differentiable with respect to u , and is unobservable but deterministic.

Theorem 6 *Given that a bivariate counting process \mathbf{N} satisfies assumptions (A1-3, 4a, 4b, 7) and is observed over $[0, T]$. Let $\hat{\lambda}_t^k$ ($k = a, b$) be the \mathcal{F}^k -conditional intensity kernel estimator of the component process N^k defined in (2.14). Assume that $M^5/T^4 \rightarrow 0$ as $T \rightarrow \infty$, $M \rightarrow \infty$ and $M/T \rightarrow 0$. Then, for any fixed $v \in [0, 1]$, the kernel estimator $\hat{\lambda}_{T_v}^k$ converges in mean squares to the conditional intensity $\lambda_{T_v}^k$, i.e.*

$$E[(\hat{\lambda}_{T_v}^k - \lambda_{T_v}^k)^2] \rightarrow 0,$$

and the normalized difference

$$\xi_v^k := \sqrt{M} \left(\frac{\hat{\lambda}_{T_v}^k - \lambda_{T_v}^k}{\sqrt{\lambda_{T_v}^k}} \right) \quad (2.18)$$

converges to a normal distribution with mean 0 and variance $\hat{\kappa}_2 = \int_{-\infty}^{\infty} \hat{K}(x) dx$, as $T \rightarrow \infty$, $M \rightarrow \infty$ and $M/T \rightarrow 0$.

By Theorem 6, it follows that $\hat{\lambda}_{T_v}^k$ is mean-squared consistent and that in the limit $\hat{\lambda}_{T_v}^k - \lambda_{T_v}^k = O_p(M^{-1/2})$ for $k = a, b$.

There is a corresponding kernel estimator of the cross-covariance function $\tilde{\gamma}_h(\cdot)$ of the innovations of the rescaled point process defined in (2.15). With an appropriate adjustment to the bandwidth, by setting the new bandwidth after rescaling H to $h = H/T$, I can reduce it to $\hat{\gamma}_H(\ell)$. For a fixed $\sigma \in [0, 1]$,

$$\begin{aligned}\hat{\gamma}_H(T\sigma) &= \frac{1}{T} \int_0^T \int_0^T K_H(t-s-T\sigma) (dN_s^a - \hat{\lambda}_s^a ds) (dN_t^b - \hat{\lambda}_t^b dt) \\ &= \frac{1}{T} \int_0^1 \int_0^1 K_H(T(v-u-\sigma)) (dN_{Tu}^a - \hat{\lambda}_{Tu}^a T du) (dN_{Tv}^b - \hat{\lambda}_{Tv}^b T dv) \\ &= \frac{T^2}{T} \int_0^1 \int_0^1 \frac{1}{H} K\left(\frac{v-u-\sigma}{H/T}\right) (d\tilde{N}_u^a - \tilde{\lambda}_u^a du) (d\tilde{N}_v^b - \tilde{\lambda}_v^b dv) \\ &= \int_0^1 \int_0^1 K_h(v-u-\sigma) d\tilde{\epsilon}_u^a d\tilde{\epsilon}_v^b =: \hat{\gamma}_h(\sigma).\end{aligned}$$

For a fixed lag $\sigma \in [0, 1]$, the kernel cross-covariance estimator $\hat{\gamma}_h(\sigma)$ consistently estimates $\tilde{\gamma}(\sigma)$ as $n^k = \tilde{N}^k(1) \rightarrow \infty$, $h \rightarrow 0$ and $n^k h \rightarrow \infty$ for $k = a, b$.

The statistic Q can thus be expressed in terms of the squared sample cross-covariance function of the rescaled point process defined in (2.15) with rescaled bandwidths. Assuming that the weighting function is another kernel with bandwidth B , i.e. $w(\ell) = w_B(\ell)$, I can rewrite Q into

$$\begin{aligned}Q &= \int_I w_B(\ell) \hat{\gamma}_H^2(\ell) d\ell \\ &= T \int_{I/T} w_B(T\sigma) \hat{\gamma}_H^2(T\sigma) d\sigma \\ &= \int_{I/T} w_b(\sigma) \hat{\gamma}_h^2(\sigma) d\sigma,\end{aligned}$$

where $b = B/T$ and $h = H/T$.

2.4.7 Simplified Statistic

Another statistic that deserves our study is

$$Q^s = \frac{1}{T^2} \int_I \int_J w_B(\ell) dN_s^a dN_{s+\ell}^b.$$

where $I \subseteq [-T, T]$ and $J = [-\ell, T - \ell] \cap [0, T]$ are the ranges of integration with respect to ℓ and s , respectively. In fact, this statistic is the continuous version of the statistic of Cox and Lewis (1972), whose asymptotic distribution was derived by Brillinger (1976). Both statistics find their root in the serial correlation statistic for univariate stationary point process (Cox, 1965). Instead of the continuous weighting function $w(\ell)$, they essentially considered a discrete set of weights on the product increments of the counting processes at a prespecified grid of lags, which are separated wide enough to guarantee the independence of the product increments when summed together.

To quantify how much we lose with the simplified statistic, let us do a comparison between Q^s and Q . If the pooled point process is simple (assumption (A1)), then the statistic Q^s is equal to, almost surely,

$$Q^s = \frac{1}{T^2} \int_I \int_J w_B(\ell) (d\hat{\epsilon}_s^a)^2 (d\hat{\epsilon}_{s+\ell}^b)^2,$$

which is the weighted integral of the squared product of residuals.²⁸ On the other hand, observe that there are two levels of smoothing in Q : the sample cross covariance $\hat{\gamma}_H(\ell)$ with kernel function $K_H(\cdot)$ which smooths the cross product increments $d\hat{\epsilon}_s^a d\hat{\epsilon}_t^b$ around the time difference $t - s = \ell$, as well as the weighting function $w_B(\ell)$ which smooths the squared sample cross-covariance function around lag $\ell = 0$. Suppose that B is large relative to H in the limit, such that

²⁸This follows from (A.5) in the Appendix.

$H = o(B)$ as $B \rightarrow \infty$. Then, the smoothing effect is dominated by $w_B(\ell)$. Indeed, as $B \rightarrow \infty$, the following approximation holds

$$w_B(\ell)K_H(t_1 - s_1 - \ell)K_H(t_2 - s_2 - \ell) = w_B(\ell)\delta_\ell(t_1 - s_1)\delta_\ell(t_2 - s_2) + o(1)$$

where $\delta_\ell(\cdot)$ is the Dirac delta function at ℓ . Hence, the difference $Q - Q^s$ becomes

$$\begin{aligned} Q - Q^s &= \int_I w_B(\ell) \hat{\gamma}_H^2(\ell) d\ell - Q^s \\ &= \frac{1}{T^2 H^2} \int_I \iiint \int_{(0,T]^4} w_B(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b d\ell - Q^s \\ &= \frac{1}{T^2} \int_I \iint_{(0,T]^4} w_B(\ell) d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{s_1+\ell}^b d\hat{\epsilon}_{s_2+\ell}^b d\ell - Q^s + o_P(1) \\ &= \frac{1}{T^2} \int_I \iint_{(0,T]^2, s_1 \neq s_2} w_B(\ell) d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{s_1+\ell}^b d\hat{\epsilon}_{s_2+\ell}^b d\ell + o_P(1). \end{aligned} \quad (2.19)$$

where in getting the second-to-last line, the quadruple integrations over $\{(s_1, s_2, t_1, t_2) \in (0, T]^4\}$ collapse to the double integrations over $\{(s_1, s_2, s_1 + \ell, s_2 + \ell) : s_1, s_2 \in (0, T]\}$.

Indeed, computing Q^s is a lot simpler than Q because there is no need to estimate conditional intensities. However, if I test the hypothesis (2.11) based on the statistic Q^s instead of Q , I will have to pay the price of potentially missing some alternatives - for example, those cases in which the cross correlations alternate in signs as the lag increases, in such a way that the integrated cross-correlation $\int_I \gamma(\ell) d\ell$ is close to zero, but the individual $\gamma(\ell)$ are not. Nevertheless, such kind of alternatives is not very common at least in our applications in default risk and high frequency finance, where the feedback from one marginal process to another is usually observed to be positively persistent, and the positive cross correlation gradually dies down as the time lag increases. In terms of computation, the statistic Q^s is much less complicated than Q since it is not necessary to estimate the sample cross covariance function $\hat{\gamma}_H(\ell)$ and the conditional intensities of the marginal processes $\hat{\lambda}_t^k$; thus two bandwidths (M and H) are saved.

The benefit of this simplification is highlighted in the simulation study where the size performance of Q^s stands out from its counterpart Q .²⁹

The mean and variance of Q^s are given in the following theorem. The techniques involved in the derivation are similar to those for Q .

Let us recall that in section 2.2, the second-order reduced form factorial product density of N^k (assumed to exist in assumption (A2)) was defined by $\varphi^{kk}(u)dtd u := E(dN_t^k dN_{t+u}^k)$ for $u \neq 0$ and $\varphi^{kk}(0)dt = E(dN_t^k)^2 = E(dN_t^k) = \lambda^k dt$. Note that there is a discontinuity point at $u = 0$ as $\lim_{u \rightarrow 0} \varphi^{kk}(u) = (\lambda^k)^2 \neq \varphi^{kk}(0)$. The reduced unconditional auto-covariance density function can then be expressed into $c^{kk}(u)dtd u := E(dN_t^k - \lambda^k dt)(dN_{t+u}^k - \lambda^k du) = [\varphi^{kk}(u) - (\lambda^k)^2]dtd u$.

Theorem 7 *Let $I \subseteq [-T, T]$ and $J_i = [-\ell_i, T - \ell_i] \cap [0, T]$ for $i = 1, 2$. Under assumptions (A1-3, 4a,b and 4d) and the null hypothesis,*

$$E(Q^s) = \frac{\lambda^a \lambda^b}{T} \int_I w_B(\ell) \left(1 - \frac{|\ell|}{T}\right) d\ell.$$

With no autocorrelations:

$$Var(Q^s) = \frac{\lambda^a \lambda^b}{T^3} \int_I w_B^2(\ell) \left(1 - \frac{|\ell|}{T}\right) d\ell.$$

²⁹There are two bandwidths for the simplified statistic: one for the weighting function and the other for the nonparametric estimator of the autocovariance function. We will show in simulations that for simple bivariate Poisson process and for bivariate point process showing mild autocorrelations, the empirical rejection rate (size) of the nonparametric test is stable over a wide range of bandwidths that satisfy the assumptions stipulated in the asymptotic theory of the statistic. When autocorrelation is high, the size is still close to the nominal level for some combinations of the bandwidths of the weighting function and the autocovariance estimators.

With autocorrelations:

$$\begin{aligned} \text{Var}(Q^s) &= \frac{1}{T^4} \iint_{I^2} \int_{J_2} \int_{J_1} w_B(\ell_1) w_B(\ell_2) c^{aa}(s_2 - s_1) c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2 \\ &\quad + \frac{(\lambda^b)^2}{T^4} \iint_{I^2} \int_{J_2} \int_{J_1} w_B(\ell_1) w_B(\ell_2) c^{aa}(s_2 - s_1) ds_1 ds_2 d\ell_1 d\ell_2 \\ &\quad + \frac{(\lambda^a)^2}{T^4} \iint_{I^2} \int_{J_2} \int_{J_1} w_B(\ell_1) w_B(\ell_2) c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2. \end{aligned}$$

If $I = [0, T]$ and $B = o(T)$ as $T \rightarrow \infty$, then (with autocorrelations)

$$\begin{aligned} \text{Var}(Q^s) &\approx \frac{2}{T^3} \left[\int_0^T \mathcal{W}_2(u) du \int_{-T}^T c^{aa}(v) c^{bb}(v + u) dv + (\lambda^b)^2 \omega_1 \int_0^T c^{aa}(v) dv \right. \\ &\quad \left. + (\lambda^a)^2 \omega_1 \int_0^T c^{bb}(v) dv \right], \end{aligned} \quad (2.20)$$

where $\omega_1 = \int_0^T w(\ell) \left(1 - \frac{\ell}{T}\right) d\ell$ and $\mathcal{W}_2(u) = \int_u^T w(\ell - u) w(\ell) \left(1 - \frac{\ell}{T}\right) d\ell$.

In practice, the mean and variance can be consistently estimated with the following replacements. For $k = a, b$:

(i) replace the unconditional intensity λ^k by the estimator $\hat{\lambda}^k = N^k/T$, and

(ii) replace the unconditional auto-covariance density $c^{kk}(\ell)$ by the kernel estimator:

$$\begin{aligned} \hat{c}_{R^k}^{kk}(\ell) &= \frac{1}{T} \int_0^T \int_0^T \ddot{K}_{R^k}(t - s - \ell) (dN_s^k - \hat{\lambda}^k ds) (dN_t^k - \hat{\lambda}^k dt) \\ &= \frac{1}{T} \sum_{i=1}^{N_T^k} \sum_{j=1}^{N_T^k} \ddot{K}_{R^k}(t_j^k - t_i^k - \ell) - \left(1 - \frac{|\ell|}{T}\right) (\hat{\lambda}^k)^2 + o(1), \end{aligned} \quad (2.21)$$

where the last equality holds if $R^k/T \rightarrow 0$ as $T \rightarrow \infty$. The proof of (2.21) will be given in Appendix A.7, which requires that $\ddot{K}(\cdot)$ satisfy the following assumption:

Assumption (A4c) The kernel function $\ddot{K}(\cdot)$ is symmetric around zero and sat-

isfies $\ddot{k}_1 \equiv \int_{-\infty}^{\infty} \ddot{K}(u) du = 1$, $\ddot{k}_2 \equiv \int_{-\infty}^{\infty} \ddot{K}^2(u) du < \infty$, $\ddot{k}_4 \equiv \iiint_{(-\infty, \infty)} \ddot{K}(u) \ddot{K}(v) \ddot{K}(u + w) \ddot{K}(v + w) du dv dw < \infty$ and $\int_{-\infty}^{\infty} u^2 \ddot{K}(u) du < \infty$.

2.5 Asymptotic Theory

2.5.1 Asymptotic Normality under the Null

Recall from the definition that the test statistic Q is the weighted integral of squared sample cross-covariance function between the residuals of the component processes. However, the residuals $d\hat{\epsilon}_t^k$ do not form a martingale difference process as the counting process increment dN_t^k is demeaned by its estimated conditional mean $\hat{\lambda}_t^k dt$ instead of the true conditional mean $\lambda_t^k dt$. According to the definition of ϵ_t^k , the innovations $d\epsilon_t^k = dN_t^k - \lambda_t^k dt$ form a martingale difference process, but not the residuals $d\hat{\epsilon}_t^k = dN_t^k - \hat{\lambda}_t^k dt$.

To facilitate the proof, it is more convenient to separate the analysis of the estimation error of conditional intensity estimators $\hat{\lambda}_t^k$ from that of asymptotic distribution of the test statistic. To this end, I define the hypothetical version of Q as follows

$$\tilde{Q} = \int_I w_B(\ell) \gamma_H^2(\ell) d\ell,$$

where $\gamma_H(\ell)$ is the hypothetical cross-covariance kernel estimator between the innovations $d\epsilon_s^a$ and $d\epsilon_t^b$:

$$\begin{aligned} \gamma_H(\ell) &= \frac{1}{T} \int_0^T \int_0^T K_H(t-s-\ell) d\epsilon_s^a d\epsilon_t^b \\ &= \frac{1}{T} \int_0^T \int_0^T K_H(t-s-\ell) (dN_s^a - \lambda_s^a ds) (dN_t^b - \lambda_t^b dt). \end{aligned}$$

In the first stage of the proof, I will prove the asymptotic normality of the hypothetical test statistic \tilde{Q} . In the second stage (to be covered in section 2.5.2), I will examine the conditions under which the approximation of \tilde{Q} by Q yields an asymptotically negligible error, so that Q is also asymptotically normally distributed.

Theorem 8 Under assumptions (A1-3,4a,5,6) and the null hypothesis (2.11), the normalized test statistic

$$J = \frac{\tilde{Q} - E(\tilde{Q})}{\sqrt{\text{Var}(\tilde{Q})}} \quad (2.22)$$

converges in distribution to a standard normal random variable as $T \rightarrow \infty$, $H \rightarrow \infty$ and $H/T \rightarrow 0$, where the mean and variance of \tilde{Q} are given as follows:

$$E(\tilde{Q}) = \frac{1}{TH} \lambda^a \lambda^b \kappa_2 \int_I w_B(\ell) \left(1 - \frac{|\ell|}{T}\right) d\ell + o\left(\frac{1}{TH}\right),$$

$$\text{Var}(\tilde{Q})$$

$$\begin{aligned} &= \frac{2}{T^2 H} \kappa_4 \int_I w_B^2(\ell) \int_{-(T-|\ell|)}^{T-|\ell|} \left(1 - \frac{|\ell|}{T} - \frac{|\ell|}{T}\right) \left[(\lambda^a)^2 + c^{aa}(r)\right] \left[(\lambda^b)^2 + c^{aa}(r)\right] dr d\ell + o\left(\frac{1}{T^2 H}\right) \\ &= \frac{2(\lambda^a \lambda^b)^2}{T^2 H} \kappa_4 \int_I w_B^2(\ell) \left(1 - \frac{|\ell|}{T}\right)^2 d\ell \\ &\quad + \frac{2}{T^2 H} \kappa_4 \int_I w_B^2(\ell) \int_{-(T-|\ell|)}^{T-|\ell|} \left(1 - \frac{|\ell|}{T} - \frac{|\ell|}{T}\right) f(r) dr d\ell + o\left(\frac{1}{T^2 H}\right), \end{aligned}$$

$$\text{where } f(x) = (\lambda^a)^2 c^{bb}(r) + (\lambda^b)^2 c^{aa}(r) dr + c^{aa}(r) c^{bb}(r).$$

If N^a and N^b do not exhibit auto-correlations, then $c^{kk}(u) \equiv 0$ for $k = a, b$ and hence the variance reduces to the first term in the last equality.

2.5.2 Effect of Estimation

In this section, I discuss the effect of estimating the unconditional and the \mathcal{F}^k -conditional intensities on the asymptotic distribution of the statistic J . I want to argue that, with the right convergence rates of the bandwidths, the asymptotic distribution of J is unaffected by both estimations.

In practice, the statistic J is infeasible because (i) \tilde{Q} is a function of the conditional intensities λ_t^a and λ_t^b ; and (ii) both $E(Q) = E(\tilde{Q})$ and $\text{Var}(Q) = \text{Var}(\tilde{Q})$

contain the unconditional intensities λ^a and λ^b . As discussed in section 2.4.4, one way to estimate the unknown conditional intensities λ_t^k (for $k = a, b$) is by means of the nonparametric kernel estimator

$$\hat{\lambda}_t^k = \int_0^T \frac{1}{M} \hat{K}\left(\frac{t-u}{M}\right) dN_u^k,$$

On the other hand, by stationarity of \mathbf{N} (assumption (A2)) the unconditional intensities λ^k (for $k = a, b$) are consistently estimated by

$$\hat{\lambda}^k = \frac{N_T^k}{T}.$$

Recall that Q is the same as \tilde{Q} after replacing λ_t^k by $\hat{\lambda}_t^k$. Let $\widehat{E}(Q)$ and $\widehat{Var}(Q)$ be the same as $E(Q)$ and $Var(Q)$ after replacing λ^k by $\hat{\lambda}^k$ and $c^{kk}(\ell)$ by $\hat{c}_{R^k}^{kk}(\ell)$.

Theorem 9 *Suppose that $H = o(M)$ as $M \rightarrow \infty$, and that $M^5/T^4 \rightarrow 0$ and $(R^k)^5/T^4 \rightarrow 0$ as $T \rightarrow \infty$. Then, under assumptions (A4b,4c) and the assumptions in Theorems 6 and 8, the statistic \hat{J} defined by*

$$\hat{J} = \frac{Q - \widehat{E}(Q)}{\sqrt{\widehat{Var}(Q)}}$$

converges in distribution to a standard normal random variable as $T \rightarrow \infty$, $H \rightarrow \infty$ and $H/T \rightarrow 0$.

As discussed in section 2.4.4, the conditional intensity λ_t^k of each component process N^k can also be modeled by a parametric model. Since the estimator of the parameter vector has the typical parametric convergence rate of $T^{-1/2}$ or $(N_T^k)^{-1/2}$ (which is faster than the nonparametric rate of $M^{-1/2}$), the asymptotic bandwidth condition in Theorem 9, i.e. $H = o(M)$ as $M \rightarrow \infty$ becomes redundant, and thus the result of Theorem 9 is still valid even without such condition. Similar remark applies to the auto-covariance density function $c^{kk}(\ell)$.

2.5.3 Asymptotic Local Power

To evaluate the local power of the Q test, I consider the following sequence of alternative hypotheses

$$\mathbf{H}_{a_T} : \gamma(\ell) = a_T \sqrt{\lambda^a \lambda^b} \rho(\ell),$$

where $a_T \rho(\ell)$ is the cross-correlation function between $d\epsilon_s^a$ and $d\epsilon_{s+\ell}^b$, and a_T is a sequence of numbers so that $a_T \rightarrow \infty$ and $a_T = o(T)$ as $T \rightarrow \infty$. The function $\rho(\ell)$, the cross-correlation function before inflated by the factor a_T , is required to be square-integrable over \mathbb{R} . The goal is to determine the correct rate a_T^* with which the test based on Q has asymptotic local power. For notational simplicity, I only discuss the case where N^a and N^b do not exhibit auto-correlations. The result corresponding to autocorrelated point processes can be stated similarly.

The following assumption is needed:

Assumption (A8) The joint cumulant $c_{22}(\ell_1, \ell_2, \ell_3)$ of $\{d\epsilon_s^a, d\epsilon_{s+\ell_1}^a, d\epsilon_{s+\ell_2}^b, d\epsilon_{s+\ell_3}^b\}$ is of order $o(a_T^2)$.

Theorem 10 Suppose that assumption (A8) and the assumptions in Theorem 8 hold. Suppose further that $H = o(B)$ as $B \rightarrow \infty$. Then, under $\mathbf{H}_{a_T^*}$ with $a_T^* = H^{1/4}$, the statistic $J - \mu(K, w_B)$ (J as defined in (A.20)) converges in distribution to $N(0, 1)$ as $H \rightarrow \infty$ and $H = o(T)$ as $T \rightarrow \infty$, where

$$\mu(K, w_B) = \frac{\kappa_2 \int_I w_B(\ell) \left(1 - \frac{|\ell|}{T}\right) \check{\rho}^2(\ell) d\ell}{\sqrt{2\kappa_4 \int_I w_B^2(\ell) \left(1 - \frac{|\ell|}{T}\right)^2 d\ell}}$$

and

$$\check{\rho}^2(\ell) := \rho^2(\ell) + \int_{-T}^T \left(1 - \frac{|u|}{T}\right) \rho\left(\ell + \frac{u}{T}\right) \rho\left(\ell - \frac{u}{T}\right) du.$$

According to Theorem 10, a test based on Q picks up equivalent asymptotic efficiency against the sequence of Pitman's alternatives in which the cross-correlation of innovations (for each lag ℓ) grows at the rate of $a_T^* = H^{1/4}$ as the sample size T tends to infinity. It is important to note, after mapping the sampling period from $[0, T]$ to $[0, 1]$ as in (2.15), that the cross-covariance under \mathbf{H}_{a_T} becomes $\tilde{\gamma}(\sigma) = \gamma(T\sigma) = a_T \sqrt{\lambda^a \lambda^b} \rho(T\sigma) = \tilde{a}_T \sqrt{\lambda^a \lambda^b} \tilde{\rho}(\sigma)$ by (2.17), where \tilde{a}_T and $\tilde{\rho}(\sigma)$ are the rate and cross-correlation of innovations after rescaling. As a result, the corresponding rate that maintains the asymptotic efficiency of the test under the new scale is $\tilde{a}_T^* = H^{1/4}/T^\nu$, where ν is the rate of decay of the uninflated cross-correlation function ρ : $\rho(\ell) = O(\ell^{-\nu})$ as $\ell \rightarrow \infty$. The rate \tilde{a}_T^* generally goes to zero for bivariate point processes exhibiting short and long memory cross-correlation dynamics, as long as $\nu \geq 1/4$.

2.6 Bandwidth Choices

According to assumption (A5), the weighting function $w(\ell)$ in the test statistic Q is required to be integrable. In practice, it is natural to choose a function that decreases with the absolute time lag $|\ell|$ to reflect the decaying economic significance of the feedback relationship over time (as discussed in section 2.4.3). Having this economic motivation in mind, I suppose in this section, without loss of generality, that the weighting function is a kernel function with bandwidth B , i.e. $w(\ell) \equiv w_B(\ell) = w(\ell/B)/B$. The bandwidth B is responsible for discounting the importance of the feedback strength as represented by the squared cross-covariance of innovations: the further away the time lag ℓ is from zero, the smaller is the weight $w_B(\ell)$.

2.6.1 Case 1: $B \ll H \ll T$

Suppose $B = o(H)$ as $H \rightarrow \infty$. This happens when B is kept fixed, or when $B \rightarrow \infty$ but $B/H \rightarrow 0$. Since $w(\ell)$ has been assumed to be a fixed function before this section, the asymptotic result in Theorem 8 remains valid. Nevertheless, I can simplify the result which is summarized in the following corollary.

Corollary 11 *Let $Q^{\mathcal{G}} \equiv \frac{TH}{\lambda^a \lambda^b} Q$. Suppose that $B = o(H)$ as $H \rightarrow \infty$. Suppose further that $I = [0, T]$. Then, with the assumptions in Theorem 8 and under the null hypothesis (2.11), the statistic*

$$\mathcal{M}^{\mathcal{G}} \equiv \frac{Q^{\mathcal{G}} - C^{\mathcal{G}}}{\sqrt{2D^{\mathcal{G}}}}$$

converges in distribution to a standard normal random variable as $T \rightarrow \infty$, and $H/T \rightarrow 0$ as $H \rightarrow \infty$, where

$$C^{\mathcal{G}} = \kappa_2$$

and

$$D^{\mathcal{G}} = 3\kappa_4.$$

2.6.2 Case 2: $H \ll B \ll T$

Suppose instead that $B \rightarrow \infty$ and $H = o(B)$ as $H \rightarrow \infty$. In this case, the smoothing behavior of the covariance estimator is dominated by that of the weighting function $w(\ell)$. As it turns out, the normalized statistic (denoted by $\mathcal{M}^{\mathcal{H}}$ in the following corollary) is equivalent to the continuous analog of Hong's (1996a) test applied to testing for cross-correlation between two time series.

Corollary 12 *Let $Q^{\mathcal{H}} \equiv \frac{TB}{\lambda^a \lambda^b} Q$. Suppose that $B \rightarrow \infty$ and that $H = o(B)$ as $H \rightarrow \infty$. Suppose further that $I = [0, T]$. Then, with the assumptions in Theorem 8 and under the null hypothesis (2.11), the statistic*

$$\mathcal{M}^{\mathcal{H}} \equiv \frac{Q^{\mathcal{H}} - C^{\mathcal{H}}}{\sqrt{2D^{\mathcal{H}}}}$$

converges in distribution to a standard normal random variable as $T \rightarrow \infty$, and $B/T \rightarrow 0$ as $B \rightarrow \infty$, where

$$C^{\mathcal{H}} = \int_0^T w\left(\frac{\ell}{B}\right) \left(1 - \frac{\ell}{T}\right) d\ell$$

and

$$D^{\mathcal{H}} = \int_0^T w^2\left(\frac{\ell}{B}\right) \left(1 - \frac{\ell}{T}\right)^2 d\ell.$$

2.6.3 Optimal Bandwidths

Choosing optimal bandwidths is an important and challenging task in nonparametric analyses. For nonparametric estimation problems, optimal bandwidths are chosen to minimize the mean squared error (MSE), and automated procedures that yield data-driven bandwidths are available and well-studied for numerous statistical models. However, optimal bandwidth selection remains a relatively unknown territory for nonparametric hypothesis testing problems. In the first in-depth analysis of how to choose the optimal bandwidth of the heteroskedasticity-autocorrelation consistent estimator for testing purpose, Sun, Phillips and Jin (2008) proposed to minimize a loss function which is a weighted average of the probabilities of type I and II error. Their theoretical comparison revealed that the bandwidth optimal for testing has a smaller asymptotic order ($O(T^{1/3})$) than the MSE-optimal bandwidth, which is typically $O(T^{1/5})$. Although the focus is on statistical inference of the simple location

model, their result could serve as a guide to the present problem of nonparametric testing for Granger causality.

2.7 Simulations

2.7.1 Size and Power of Q

In the first set of size experiments, the data generating process (DGP) is set to be a bivariate Poisson process which consists of two independent marginal Poisson processes with rate 0.1. The number of simulation runs is 5000. The weighting function of Q is chosen to be a Gaussian kernel with bandwidth $B = 10$. I consider four different sample lengths ($T = 500, 1000, 1500, 2000$) with corresponding bandwidths ($M = 60, 75, 100, 120$) for the nonparametric conditional intensity estimators in such a way that the ratio M/T gradually diminishes. Figure 2 shows the plots of the empirical rejection rates against different bandwidths H of the sample cross-covariance estimator for the four different sample lengths we considered. The simulation result reveals that in finite sample the test is generally undersized at the 0.1 nominal level and oversized at the 0.05 nominal level, but the performance improves with sample length.

In the second set of experiments, the DGP is set to a more realistic one: a bivariate exponential Hawkes model (see section 2.1.4) with parameters

$$\mu = \begin{pmatrix} 0.0277 \\ 0.0512 \end{pmatrix}, \alpha = \begin{pmatrix} 0.0086 & 0.0017 \\ 0 \text{ or } 0.0182 & 0.0896 \end{pmatrix}, \beta = \begin{pmatrix} 0.0254 & 0.0507 \\ 0.0254 & 0.1473 \end{pmatrix}, \quad (2.23)$$

which were estimated by fitting the model to a high frequency TAQ dataset of PG traded in NYSE on a randomly chosen day (1997/8/8) and period (9:45am

to 10:15am). For the size experiments, the parameter α_{21} was intentionally set to zero so that there is no causal relation from the first process to the second under the DGP, and we are interested in testing the existence of causality from the first process to the second only (i.e. by setting the integration range of the statistic Q to $I = [0, T]$). The number of simulation runs is 10000 with a fixed sample length 1800 (in seconds). The bandwidth of the sample cross covariance estimator is fixed at $H = 3$. A Gaussian kernel with bandwidths $B = 2$ and 20 respectively is chosen for the weighting function. For the power experiments, I set α_{21} back to the original estimate 0.0182. There is an increase, albeit mild, in the rejection rate compared to the size experiments. Figure 3 shows the plots of the rejection rates against different bandwidths M of the nonparametric conditional intensity estimators. A first observation, after comparing Figures 3(a) and 3(c), is that the empirical sizes of the test are more stable over various M when B is small. A second observation, after comparing Figures 3(b) and 3(d), is that the test seems to be more powerful when B is small. This indicates that, while a more slowly decaying weighting function gives a more consistent test against alternatives with longer causal lags, this is done at the expense of a lower power and more sensitive size to bandwidth choices.

2.7.2 Size and Power of Q^s

To investigate the finite sample performance of the simplified statistic Q^s , I conduct four size experiments with different parameter combinations of a bivariate exponential Hawkes model. Recall that there are only three bandwidths to choose for Q^s , namely the bandwidth B of the weighting function $w_B(\ell)$ and the bandwidths R^k of the autocorrelation function estimator $\hat{c}_{R^k}^{kk}(\ell)$ for $k = a, b$. In

each of the following experiments, I generate four sets of 5000 samples of various sizes ($T = 300, 600, 900, 1200$) from a DGP and carry out a Q^s test for Granger causality from N^a to N^b on each of the samples. The DGP's of the four size experiments and one power experiment are all bivariate exponential Hawkes models with the following features:

- Size experiment 1: N^a and N^b are independent and have the same unconditional intensities with comparable and moderate self-excitatory (autoregressive) strength (Figure 4).
- Size experiment 2: N^a and N^b are independent and have the same unconditional intensities, but N^b exhibits stronger self-excitation than N^a (Figure 5).
- Size experiment 3: N^b Granger causes N^a , and both have the same unconditional intensities and self-excitatory strength (Figure 6).
- Size experiment 4: N^a and N^b are independent and have the same self-excitatory strength, but unconditional intensity of N^b doubles that of N^a (Figure 7).
- Size experiment 5: N^a and N^b are independent and have the same unconditional intensities with comparable and highly persistent self-excitatory (autoregressive) strength (Figure 8).
- Power experiment: N^a Granger causes N^b , and both have the same unconditional intensities and self-excitatory strength (Figure 9).

The nominal rejection rates are plotted against different weighting function bandwidths B (small relative to T). The bandwidths R^k of the autocovariance

function estimators are set proportional to B ($R^k = cB$ where $c = 0.5$ for size experiment 5 and $c = 1$ for all other experiments).

Under those DGP's that satisfies the null hypothesis of no Granger causality (all size experiments), the empirical rejection rates of the test based on Q^s are reasonably close to the nominal rates over a certain range of B that grows with T , as shown in Figures 4-8. According to Theorem 7, I need $B = o(T)$ so that the variance can be computed by (2.20) in the theorem. In general, the empirical size becomes more accurate as the sample length T increases. On the other hand, the Q^s test is powerful against the alternative of a bivariate exponential Hawkes model exhibiting Granger causality from N^a to N^b , and the power increases with sample length T , as shown in Figure 9.

To study the test performance under more realistic settings, I set the DGPs to some bivariate ACI(1,1) models estimated from bankruptcy contagion data over different periods. The models are estimated in Chapter 3 and the results are displayed in Tables B.1 and B.2. The estimated models reveal some traces of non-stationarity in the data (as the persistence parameters sum up to over one in magnitude). The data consist of the bankruptcy times of upstream (group U)³⁰ and downstream (group D)³¹ over two sampling periods. In the first period, from March 2001 to Nov 2001, the estimated model (see Table B.1) reveals a stronger bankruptcy contagion spreading from downstream to upstream ($\hat{\beta}^{ab} = 1.91$, $\hat{\alpha}^{ba} = -17.84$) than in the opposite direction $\hat{\beta}^{ba} = 0.08$, $\hat{\alpha}^{ab} = 0.51$). Therefore, the nonparametric Q^s test for the direction $U \rightarrow D$ ($D \rightarrow U$) approximately constitutes a size (power) experiment. The story is opposite in the second period (from July 2009 to November 2011), in which the estimated model

³⁰SIC division codes: A, B, C, D and E.

³¹SIC division codes: F, G, H and I.

(see Table B.2) reveals a mild bankruptcy contagion spreading from upstream to downstream ($\hat{\beta}^{ba} = 1.01$, $\hat{\alpha}^{ab} = -25.42$), and small feedback in the opposite direction ($\hat{\beta}^{ab} = 0.02$, $\hat{\alpha}^{ba} = 0.74$). Hence, the nonparametric Q^s test for the direction $D \rightarrow U$ ($U \rightarrow D$) approximately constitutes a size (power) experiment.

In summary, we have the following size and power experiments:

- Size experiment R-S1: Q^s test for the direction $U \rightarrow D$. DGP: estimated bivariate ACI(1,1) model over March 2001 to November 2001.
- Size experiment R-S2: Q^s test for the direction $D \rightarrow U$. DGP: estimated bivariate ACI(1,1) model over July 2009 to November 2011.
- Power experiment R-P1: Q^s test for the direction $D \rightarrow U$. DGP: estimated bivariate ACI(1,1) model over March 2001 to November 2001.
- Power experiment R-P2: Q^s test for the direction $U \rightarrow D$. DGP: estimated bivariate ACI(1,1) model over July 2009 to November 2011.

The simulation results are presented from Figures A.10 to A.13, in which the empirical rejection rates are plotted against the ratio of the weighting function bandwidth to the sample length, B/T . In the size experiments, the rejection rates of the Q^s test are reasonably close to the nominal levels over a range of the B/T ratios that are consistent with the rates governed by the asymptotic theory.³² The size of the test becomes more accurate while the power grows as the sample size $N(T)$ increases. The size and power performance seems to be robust to the non-stationary nature of the DGPs under study, providing support to the empirical applications of the nonparametric Q^s test in Chapter 3.

³²The inflated rejection rates for small B/T in the size experiment R-S2 may be contributed by the short-lived feedback from the innovation of downstream bankruptcy point process to the upstream process ($\hat{\alpha}^{ba} = 0.74$).

2.8 Applications

2.8.1 Trades and Quotes

In the market microstructure literature, there are various theories that attempt to explain the trades and quotes dynamics of stocks traded in stock exchanges. In the seminal study, Diamond and Verrecchia (1987) propose that the speed of price adjustment can be asymmetric due to short sale constraints. As a result, a lack of trades signals bad news because informed traders cannot leverage on their insights and short-sell the stock. Alternatively, Easley and O'hara (1992) argue that trade arrival is related to the existence of new information. Trade arrival affects the belief on the fundamental stock price held by dealers, who learn about the direction of new information from the observed trade sequence and adjust their bid and/or ask quotes in a Bayesian manner. It is believed that a high trade intensity is followed by more quote revisions, while a low trade intensity means a lack of new information transmitted to the market and hence leads to fewer quote revisions. As discussed in 2.1.3, much existing research is devoted to the testing of these market microstructure hypotheses, but the tests are generally conducted through statistical inference under strong parametric assumptions (e.g. VAR model in Hasbrouck, 1991 and Dufour and Engle, 2000; the bivariate duration model in Engle and Lunde, 2003). This problem offers an interesting opportunity to apply the nonparametric test in this chapter. With minimal assumptions on the trade and quote revision dynamics, the following empirical results indicate the direction and strength of causal effect in support of the conjecture of Easley and O'hara (1992): more trade arrivals predict more quote revisions.

I obtain the data from TAQ database available in the Wharton Research Data Services. The dataset consists of all the transaction and quote revision timestamps of the stocks of Proctor and Gamble (NYSE:PG) in the 41 trading days from 1997/8/4 to 1997/9/30, the same time span as the dataset of Engle and Lunde (2003). Then, following the standard data cleaning procedures (e.g. Engle and Russell, 1998) to prepare the dataset for further analyses,

1. I employ the five-second rule when combining the transaction and quote time sequences into a bivariate point process by adding five seconds to all the recorded quote timestamps. This is to reduce unwanted effects from the fact that transactions were usually recorded with a time delay.
2. I eliminate all transaction and quote records before 9:45am on every trading day. Stock trades in the opening period of a trading day are generated from open auctions and are thus believed to follow different dynamics.
3. Since the TAQ timestamps are accurate up to a second, this introduces a limitation to the causal inference in that there is no way to tell the causal direction among those events happening within the same second. The sampled data also constitutes a violation of assumption (A1). I treat multiple trades and quotes occurring at the same second as one event, so that an event actually indicates the occurrence of at least one event within the same second.³³

³³For PG, 5.6% of trades, 28.1% of quote revisions and 3.6% of trades and quotes were recorded with identical timestamps (in seconds). The corresponding proportions for GM are 5.7%, 19.9% and 2.6%, respectively. Admittedly, the exceedingly number of quote revisions recorded at the same time invalidates assumption (A1), but given the low proportions for trades and trade-quote pairs with same timestamps, the distortion to the empirical results is on the conservative side. That is, if there exists a more sophisticated Granger causality test that takes into account the possibility of simultaneous quote events, the support for trade-to-quote causality would be even stronger than the support Q and Q^s tests provide, as we shall see later in Tables A.2-A.6.

After carrying out the data cleaning procedures, I split the data into different trading periods and conduct the nonparametric causality test for each trading day. Then, I count the number of trading days with significant causality from trade to quote (or quote to trade) dynamics. For each sampling period, let N^t and N^q be the counting processes of trade and quote revisions, respectively. The hypotheses of interest are

\mathbf{H}_0 : there is no Granger causality from N^a to N^b ; vs

\mathbf{H}_1 : there is Granger causality from N^a to N^b .

where $a, b \in \{t, q\}$ and $a \neq b$.

The results are summarized in Tables A.2 to A.4. In each case, I present the significant day count for different combinations of bandwidths (all in seconds). For each (H, B) pair, the bandwidth M of the conditional intensity estimator is determined from simulations so that the rejection rate matches the nominal size.

Some key observations are in order. First, there are more days with significant causation from trade to quote update dynamics than from quote update to trade dynamics for most bandwidth combinations. This supports the findings of Engle and Lunde (2003). Second, for most bandwidth combinations, there are more days with significant causations (in either direction) during the middle of a trading day (11:45am – 12:45pm) than in the opening and closing trading periods (9:45am – 10:15am and 3:30pm – 4:00pm). One possible explanation is that there are more confounding factors (e.g. news arrival, trading strategies) that trigger a quote revision around the time when the market opens and closes. When the market is relatively quiet, investors have less sources to rely on but update their belief on the fundamental stock price by observing the recent transactions. Third, the contrast between the two causation directions be-

comes sharper in general when the weighting function, a Gaussian kernel, decays more slowly (larger B), and it becomes the sharpest in most cases when B is 10 seconds (when the day counts with significant causation from trade to quote is the maximum). This may suggest that most causal dynamics from trades to quotes occur and finish over a time span of about $3B = 30$ seconds.

Next, I employ the simplified statistic Q^s to test the data. I am interested to see whether it implies the same causal relation from trades to quotes as found earlier, given that a test based on Q^s is only consistent against a smaller set of alternatives (as discussed in section 2.4.7). The result of the Q^s test on trade and quote revision sequences of PG is presented in Table A.5. The result shows stronger support for the causal direction from trades to quote revisions across various trading periods of a day (compare Table A.5 to Tables A.2-A.4: the Q^s test uncovers more significant days with trade-to-quote causality than the Q test does). I also conduct the Q^s test on trades and quotes of GM, and obtain similar result that trades Granger-cause quote revisions. (See Table A.6 for the test results on General Motors. Test results of other stocks considered by Engle and Lunde (2003) are similar and available upon request.) The stronger support by the Q^s test for the trade-to-quote causality suggests indirectly that the actual feedback resulting from a shock in trade dynamics to quote revision dynamics is persistent rather than alternating in signs over the time range covered by the weighting function $w(\ell)$. Given that I am testing against the alternatives with persistent feedback effect from trades to quote revisions, it is natural that the Q test is less powerful than the Q^s test. This is the price for being consistent against a wider set of alternatives³⁴.

³⁴This includes those alternatives in which excitatory and inhibitory feedback effect from trades to quotes alternate as time lag increases.

2.8.2 Credit Contagion

Credit contagion occurs when a credit event (e.g. default, bankruptcy) of a firm leads to a cascade of credit events of other firms (see, for example, Jorion and Zhang, 2009). This phenomenon is manifested as a cluster of firm failures in a short time period. As discussed in section 2.1.4, a number of reduced-form models, including conditional independence and self-exciting models, are available to explain the dependence of these credit events over time, with varying level of success. Conditional independence model assumes that the probabilities of a credit events of a cross section of firms depend on some observed common factors (Das, Duffie, Kapadia and Saita, 2008; DDKS hereafter). This modeling approach easily induces cross-sectional dependence among firms, but is often inadequate to explain all the observed clustering of credit events unless a good set of common factors is discovered. One way to mitigate the model inadequacy is to introduce latent factors into the model (Duffie, Eckners, Horel and Saita, 2010; DEHS hereafter). Counterparty risk model, on the other hand, offers an appealing alternative: the occurrence of credit events of firms are directly dependent on each other (Jarrow and Yu, 2001). This approach captures directly the mutual-excitatory (or serial correlation) nature of credit events that is neglected by the cross-sectional approach of conditional independence models. In a series of empirical studies, Jorion and Zhang (2007, 2009) provided the first evidence that a significant channel of credit contagion is through counterparty risk exposure. The rationale behind their arguments is that the failure of a firm can affect the financial health of other firms which have business ties to the failing firm. This empirical evidence highlights the importance of counterparty risk model as an indispensable tool for credit contagion analysis.

All the aforementioned credit risk models cannot avoid the imposition of ad-hoc parametric assumptions which are not justified by any structural models. For instance, the conditional independence models of DDKS and DEHS rely on strong log-linear assumption³⁵ on default probabilities, while the counterparty risk model of Jarrow and Yu adopt a convenient linear configuration. Also, the empirical work of Jorion and Zhang is based on the linear regression model. The conclusions drawn from these parametric models have to be interpreted with care as they may be sensitive to the model assumptions. Indeed, as warned by DDKS, a rejection of their model in goodness-of-fit tests can indicate either a wrong log-linear model specification or an incorrect conditional independence assumption of the default intensities, and it is impossible to distinguish between them from their test results. Hence, it is intriguing to investigate the extent of credit contagion with as few interference from model assumptions as possible. The nonparametric Granger causality tests make this model-free investigation a reality.

I use the “Bankruptcies of U.S. firms, 1980–2010” dataset to study credit contagion. The dataset is maintained by Professor Lynn LoPucki of UCLA School of Law. The dataset records, among other entries, the filing dates of Chapter 11 and the Standard Industrial Classification (SIC) codes of big bankrupting firms³⁶. In this analysis, a credit event is defined as the occurrence of bankruptcy event(s). To be consistent with assumption (A1), I treat multiple bankruptcies on the same date as one bankruptcy event. Figure 14 shows the histogram of bankruptcy oc-

³⁵In the appendix of their paper, DEHS evaluates the robustness of their conclusion by considering the marginal nonlinear dependence of default probabilities on the distance-to-default. Nevertheless, the default probability is still assumed to link to other common factors in a log-linear fashion.

³⁶The database includes those debtor firms with assets worth \$100 million or more at the time of Chapter 11 filing (measured in 1980 dollars) and which are required to file 10-ks with the SEC. However, it may have excluded some interesting bankruptcies.

currences in 1980–2010.

I classify the bankrupting firms according to the industrial sector. More specifically, I assume that a bankruptcy belongs to manufacturing related sectors if the SIC code of the bankrupting firm is from A to E, and financial related sectors if the SIC code is from F to I. The rationale behind the classification is that the two industrial groups represent firms at the top and bottom of a typical supply chain, respectively. The manufacturing related sectors consist of agricultural, mining, construction, manufacturing, transportation, communications and utility companies, while the financial related sectors consist of wholeselling, retailing, financial, insurance, real estate and service provision companies.³⁷ Let N^m and N^f be the counting processes of bankruptcies from manufacturing and financial related sectors, respectively. Figure 15 plots the counting processes of the two types of bankruptcies. The hypotheses of interest are

\mathbf{H}_0 : there is no Granger causality from N^a to N^b ; vs

\mathbf{H}_1 : there is Granger causality from N^a to N^b .

where $a, b \in \{m, f\}$ and $a \neq b$.

Similar to the TAQ application, I carry out the Q test for different combinations of bandwidths (in days). The bandwidths M (for conditional intensity estimators) and B (for weighting function) are set equal to 365, 548 and 730 days (corresponding to 1, 1.5 and 2 years), while the bandwidth H (for cross-covariance estimator) ranges from 2 to 14 days.³⁸ The test results are displayed in Tables A.7-A.9. For most bandwidth combinations, the Q test detects a signif-

³⁷The industrial composition of bankruptcies in manufacturing related sectors are A: 0.2%; B: 6.3%; C: 4.5%; D: 58.6%; E: 30.5%. The composition in financial related sectors are F: 8.4%; G: 29.4%; H: 32.8%; I: 29.4%.

³⁸It turns out that the Q test is more sensitive to the choice of M than B . The choice of bandwidth H is guided by the restriction $H = o(M)$ from Theorem 9.

icant credit contagion (at 5% significance level) from financial to manufacturing related sectors in periods that contain crises and recession (Asian financial crisis and 9/11 in September 1996 – July 2003; subprime mortgage crisis in September 2007 – June 2010) but not in periods of economic growth (August 2003 – August 2007). The reverse contagion becomes statistically significant too during the subprime mortgage crisis.

I also conduct the Q^s test over the period September 1996 – June 2010 that spans the financial crises and the boom in the middle. During this period, there are 350 and 247 bankruptcies in the manufacturing and financial related sectors. The normalized test statistic values (together with p-values) are presented in Table A.10. The bandwidth B of the weighting function ranges from 30 to 300 days, while the bandwidths R^k of the unconditional autocorrelation kernel estimators $\hat{c}_{R^k}^{kk}(\ell)$ (for $k = m$ and f) are both fixed at 300 days. All kernels involved are chosen to be Gaussian. Over the period of interest, there is significant (at 5% significance level) credit contagion in both directions up to $B = 90$ days, but the financial-to-manufacturing contagion dominates manufacturing-to-financial contagion in the long run.

2.8.3 International Financial Contagion

The Granger causality test developed in this chapter can be used to uncover financial contagion that spreads across international stock markets. An adverse shock felt by one financial market (as reflected by very negative stock returns) often propagates quickly to other markets in a contagious manner. There is no agreement in the concept of financial contagion in the literature³⁹. For in-

³⁹See Forbes and Rigobon (2002) and the references therein for a literature review.

stance, Forbes and Rigobon (2002; hereafter FR) defined financial contagion as a significant increase in cross-market linkages after a shock. To measure and compare the extent of contagion over different stock market pairs, FR used a bias-corrected cross-correlation statistic for index returns. However, whether the increased cross-correlation represents a causal relationship (in Granger sense) is unclear. More recently, Aït-Sahalia, Cacho-Diaz and Laeven (2010; hereafter ACL) provided evidence of financial contagion by estimating a parametric Hawkes jump-diffusion model to a cross-section of index returns. The contagion concept ACL adopted is in a wider sense than that of FR in that contagion can take place in both “good” and “bad” times (see footnote 2 of ACL). Based on the dynamic model of ACL, it is possible to infer the causal direction of contagion from one market to another. Nevertheless, their reduced-form Hawkes jump-diffusion model imposes a fair amount of structure on both the auto- and cross-correlation dynamics of the jumps of index returns without any guidance from structural models. The conclusion drawn from ACL regarding causal direction of contagion is model-specific and, even if the model is correct, sensitive to model estimation error. To robustify the conclusion, it is preferred to test for Granger causality of shock propagations in a nonparametric manner.

To this end, I collect daily values of the major market indices from finance.yahoo.com and compute daily log-returns from adjusted closing values. The indices in my data are picked from representative stock markets worldwide covering various time zones, including the American (Dow Jones), European (FTSE, DAX, CAC 40), Asian Pacific (Hang Seng, Straits Times, Taiwan, Nikkei), and Australian (All Ordinary) regions. The data frame, trading hours and number of observations are summarized in Table A.11.

To define the days with negative shocks, I use the empirical 90%, 95% and 99% value-at-risk (VaR) for the corresponding stock indices. An event is defined as a negative shock when the daily return exceeds the VaR return. In each test, I pair up two point processes of events from two indices of different time zone, with a sampling period equal to the shorter of the two sample lengths of the two indices. The event timestamps are adjusted by the time difference between the two time zones of the two markets. Define the counting processes of shock events for indices a and b by N^a and N^b , respectively. The hypotheses of interest are

\mathbf{H}_0 : there is no Granger causality from N^a to N^b ; vs

\mathbf{H}_1 : there is Granger causality from N^a to N^b .

The results of the Q^s test applied to the pairs HSI-DJI, NIK-DJI, FTSE-DJI and AOI-DJI are shown in Tables A.12-A.15⁴⁰. There are a few observations. First, days with extreme negative returns exceeding 99% VaR have a much stronger contagious effect than those days with less negative returns (exceeding 95% or 90% VaR). This phenomenon is commonly found for all pairs of markets. Second, except for European stock indices, the U.S. stock market, as represented by DJI, plays a dominant role in infecting other major international stock markets. It is not hard to understand why the daily returns of European stock indices (FTSE, DAX, CAC 40) Granger-cause DJI's daily returns given the overlap of the trading hours of European stock markets and the U.S. stock market. Nonetheless, the causality from the American to European markets remains significant (for $B \leq 3$ with 95% VaR as the cutoff). Third, the test statistic values

⁴⁰The Q^s test results for pairs involving DAX and CAC are qualitatively the same as that involving FTSE (all of them are in the European time zones), while the test results for pairs involving STI and TWI are qualitatively the same as that involving HSI (all of them are in the same Asian-Pacific time zone). I do not present these results here to reserve space, but they are available upon request.

are pretty stable over different choices of B and R^k ($k = a, b$). I used different functions of $R^k = M(B)$, such as a constant $R^k = 10$ and $R^k = 24B^{0.25}$, and found that, qualitatively, the dominating indices / markets remain the same as before (when $R^k = 10.5B^{0.3}$). Fourth, the shorter is the testing window (bandwidth B of the weighting function $w(\ell)$), the stronger is the contagious effect. For instance, with 95% value-at-risk as cutoff, DJI has significant Granger causality to HSI and NIK when $B \leq 3$ (in days) and to AOI when $B \leq 5$. This implies that contagious effect, once it starts, is most significant on the first few days, but usually dampens quickly within a week.

2.9 Conclusion

With growing availability of multivariate high frequency and/or irregularly spaced point process data in economics and finance, it becomes more and more of a challenge to examine the predictive relationship among the component processes of the system. One important example of such relationship is Granger causality. Most of the existing tests for Granger causality in the traditional discrete time series setting are inadequate for the irregularity of these data. Tests based on parametric continuous time models can better preserve the salient features of the data, but they often impose strong and questionable parametric assumptions (e.g. conditional independence as in doubly stochastic models, constant feedback effect as in Hawkes models) that are seldom supported by economic theories and, more seriously, distort the test results. This calls for a need to test for Granger causality (i) in a continuous time framework and (ii) without strong parametric assumptions. In this chapter, I study a nonparametric approach to Granger causality testing on a continuous time bivariate point

process that satisfies mild assumptions. The test enjoys asymptotic normality under the null hypothesis of no Granger causality, is consistent, and exhibits nontrivial power against departure from the null. It performs reasonably well in simulation experiments and shows its usefulness in three empirical applications: market microstructure hypothesis testing, checking the existence of credit contagion between different industrial sectors, and testing for financial contagion across international stock exchanges.

In the first application on the study of market microstructure hypotheses, the test confirms the existence of a significant causal relationship from the dynamics of trades to quote revisions in high frequency financial datasets. The next application on credit contagion reveals that U.S. corporate bankruptcies in financial related sectors Granger-cause those in manufacturing related sectors during crises and recessions. Lastly, the test is applied to study the extent to which an extreme negative shock of a major stock index transmits across international financial markets. The test confirms the presence of contagion, with U.S. and European stock indices being the major sources of contagion.

CHAPTER 3

**CREDIT CONTAGION FROM WALL STREET TO MAIN STREET: AN
EMPIRICAL STUDY OF US CORPORATE BANKRUPTCIES**

3.1 introduction

The recent subprime mortgage crisis originated from the financial and mortgage sector, but its disastrous effect rippled across almost all other sectors of the economy at an unprecedented scale. From the points of view of policy makers, regulators and portfolio investors, it is paramount to understand the causes and mechanisms of credit contagion across different sectors of an economy.

There are a number of channels through which credit risk transmits from one sector of the economy to another. They may be classified into two categories:

1. Credit risk may spread through information channels. Under this hypothesis, a negative shock in one sector (in the form of a stock price decline or a bankruptcy of a firm in that sector) reveals the weak fundamentals of that sector or the whole economy. Through Bayesian learning, economic agents adjust their beliefs on the prospect of other sectors accordingly.
2. Credit risk may spread through direct interconnections (or business ties) among firms in different sectors of an economy. The interconnections constitute a complex network of firms which are bound by contractual agreements. Some examples include a supply chain and a line of credit with a bank extending loans to a business owner. A natural consequence is that all parties in the same network are exposed to some counterparty risk. When one party fails to meet the obligation in the contractual agreement,

the other party bears the brunt by suffering a financial loss or even going default or bankrupt. As a result, counterparty risk spreads across the interconnected business network that may span multiple sectors.

Both channels of contagion are well studied from the theoretical points of view. The first channel (information contagion) is illustrated by Acharya and Yorulmazer (2008) in the context of a bank run: the bad performance of a bank conveys negative signals of unfavorable loan returns in the banking industry to risk-averse depositors, who will update the belief and require a higher rate of return from other banks. The second channel (contagion by direct interconnections) is motivated by the seminar work of Jarrow and Yu (2001) on the modeling of counterparty risk. Allen and Gale (2000) explain how financial contagion naturally occurs in equilibrium and how it spreads from one bank to another.

There has been no consensus on the definition of contagion. In the literature of international financial contagion, Forbes and Rigobon (2002) give a definition in terms of increased co-movement of stock markets, although it is not clear how to pin down on the sources of contagion or at least the direction in which contagion spreads. In general, there is no hope to identify the direction of contagion without using frequently sampled data. More recently, using daily stock market indices, Ait-Sahalia, et al. (2011) are able to unveil the direction of financial contagion among major international financial markets. Their jump diffusion model contains a jump component characterized by a Hawkes process, and can be estimated by GMM. There are some limitations to the model, however. First, the conclusion of the test is sensitive to the parametric model specification. Second, even though the model works well for stationary data, the results of inference are questionable when the data exhibit non-stationarity,

a common feature for data that span across periods with contagion.

The direction and intensity of contagion transmission during financial crises and economic recessions provide useful hints for identifying the source and severity of the underlying contagion. With well-defined concepts of credit contagion and economic sectors, we are able to measure and analyze the phenomenon objectively. To this end, we define a credit contagion as follows:

Definition 13 *A credit contagion from sector A to sector B occurs if past occurrences of credit events in sector A Granger cause future occurrences of credit events in sector B.*

Definition 14 *A credit contagion in sector A occurs if past occurrences of credit events in sector A Granger cause future occurrences of credit events in sector A.*

In general, a credit event can be a default, a bankruptcy (Chapter 11 filing). In the empirical analysis of this chapter, we will focus on bankruptcy contagion in which the credit events are solely Chapter 11 filings.¹ Specifically, we are interested in whether and how the bankruptcies in one sector Granger cause the bankruptcies in another sector.

Two remarks regarding the empirical study of bankruptcy contagion are in order. First, to identify the direction of contagion, we employ a fully continuous time set-up to analyze finely collected bankruptcy times. In contrast to the monthly counts of credit events used by many previous research (Das, et al.,

¹Although Chapter 11 filings are often preceded by other indicators such as significant stock price decline and sharp increase in the price of credit derivatives, the dates on which they occur are objectively available. There is no evidence that a Chapter 11 filing in one sector takes less time than those in another sector.

2007; Duffie et al., 2009), the bankruptcy times in our dataset are accurate up to a day. The time difference between bankruptcy occurrences contains valuable information of Granger causal dynamics, which in turn provides evidence for inferring the direction and relative magnitude of bankruptcy contagion.

Second, to get rid of the unwanted effect of parametric assumptions on inference, we nonparametrically estimate the marginal conditional intensities that capture the bankruptcy contagion within a sector. The magnitude of bankruptcy contagion across two sectors can be measured by the cross correlation of the residuals obtained after filtering away the same-sector contagion from the observed bankruptcy sequences. It can be shown that the cross correlation function of residuals fully characterizes the Granger causality between any pair of bankruptcies originating from both sectors (see Chapter two).

There are two families of credit risk models:

1. Common factor models. Some examples are Cox's doubly stochastic model (i.e. conditional independence model) as in Das, et al. (2007) and frailty model as in Duffie, et al. (2009).
2. Counterparty risk models. Some examples include the mutually exciting model of Jarrow and Yu (2001) and top-down cluster model as in Azizpour, et al. (2011).

There are two kinds of directions in which a credit contagion may spread:

1. Contagion may spread among firms within the same industry (lateral direction). It may spread through information channel. For example, firms in the same industry are exposed to similar risks or common factors. It

may transmit through counterparty risk exposure as well: e.g. interbank loans.

2. Contagion may spread across different industries (longitudinal direction).
If there is a lack of demand (supply), the supply (demand) side will suffer. Example 1 (supply driven contagion): the bankruptcy of CIT group on November 1, 2009 led to a drop in the supply of affordable loans to small to mid-sized businesses. The retailers, factories and other business owners who previously relied on CIT's lines of credit would have a hard time getting affordable financing solutions (see the NY Times report "CIT's Troubles Are Small Business's Troubles" on July 13, 2009). As a result, their liquidity position was at risk, and they might be forced to go bankrupt. Example 2(demand driven contagion): In economic recessions, demands for goods and services drop. As a result, retailers and services providers (e.g. restaurant chains, shops, department stores) suffer loss and close down. They are followed by the failure of manufacturers and then raw material suppliers.

There exists related theoretical work on the causes and mechanisms of credit contagion. Drawing upon the network theory, Battiston, et al. (2007) provided a theoretical explanation of the domino effect of bankruptcy. The bankruptcy of one firm increases the probability of bankruptcy of other closely connected firms through production and/or credit relationships, and hence may eventually trigger an avalanche of bankruptcies of firms along a supply chain and/or over a counterparty network. In fact, liquidity shortage and deterioration is a noteworthy source of credit contagion. He and Xiong (2012) argue theoretically that firms may suffer from increased "credit risk" if the debt market liquidity deteriorates. A common theme reverberates in all of the above work: credit

contagion exists naturally in an incomplete market in a rational setting.

Empirical studies on credit contagion across sectors are scarce. Hertz, et al (2008) found empirical evidence that bankruptcy filing of a firm can have significant wealth effect (in the form of abnormal stock returns) on not just its competitors in the same industry but also its suppliers/customers on the same supply chain.

Das, et al. (2007) reject their proposed doubly stochastic model and admit that their model cannot explain all the observed default clustering in the data. Indeed, their test is a joint test of the specification of the default intensity and the conditional independence assumption, so a rejection can be due to the wrong log-linear functional form, the missing of important covariates, apart from the conditional independence assumption. In fact, Lando and Nielsen (2011) fail to reject the doubly stochastic model after including an additional covariate (industrial productivity index). Duffie, et al. (2009) propose a frailty model that includes unobserved factors, and it boosts up the explanatory and predictive power of its predecessor in Das, et al. Azizpour, et al. (2011) estimate a mixed cluster-frailty model and show that both the self-exciting and frailty aspects are important for explaining the stylized fact.

In all the above empirical work, the statistical inference is valid only up to the assumed parametric model. No questions were raised on the validity of the model. A more serious question, from the statistical point of view, is that there is no way to test empirically whether a conditional independence model or a cluster model is the true DGP. Barlett (1964) shows that both the doubly stochastic model and the cluster model can generate identical data. Suppose the DGP of a sequence of bankruptcy occurrences is a doubly stochastic model

with a significant common factor. If the modeler proposed to estimate a cluster model for the data, the fitted model would reveal a significant clustering effect. It would be incorrect for the modeler to infer from the estimated model that clustering effect dominates.

This is why it is important to have a model-free way to characterize credit contagion. It turns out that it is meaningful to discuss the directions of credit contagion. We will see that the tools developed in Chapter 2 are indispensable for this purpose.

Traditionally, VAR models provide a popular parametric framework to test for Granger causality, but it relies on fixed sampling intervals. However, bankruptcies occur irregularly over time. Inferring from an estimated VAR model on bankruptcy count data may lead to erroneous conclusion. In particular, spurious causality and/or non-causality may occur when the time duration in which the causal dynamics is transmitted is shorter than the sampling interval. In section 3.2, we will show that VAR model is not a suitable model for testing for Granger causality for irregularly spaced bankruptcy times. For different choices of sampling intervals, VAR models can give contradictory results on the feedback effect from one counting process to another. This is attributable to the temporal aggregation of random event data when forming the time series of counts with fixed sampling intervals that are too wide relative to the actual causal durations. In general, any parametric time series models based on such coarsely sampled count sequences would inevitably yield unreliable statistical inference regarding the Granger causal dynamics of the original point process, as there is a loss of information during the discretization process. Even though a parametric multivariate Poisson autoregressive model on count sequences (e.g.

Heinen and Rengifo, 2007) surpasses VAR models in that it is designed for the count data, they would still fail to capture the causal dynamics if the sampling frequency is not high enough.

This highlights the remarks by Sims (1971) and Granger (1988) which suggest that there is no hope of disentangling the Granger causal directions unless the sampling intervals are shrunk to be shorter than the Granger causal duration. However, the highly irregular and random nature of event times implies that no discrete time series model with fixed-width sampling intervals would pick up all Granger causal dynamics while still maintaining parsimony. There are at least two approaches that can circumvent the problem: parametric multivariate intensity modeling; and nonparametric Granger causality testing on point process data.

A prominent example of the first approach is represented by the Multivariate Autoregressive Conditional Intensity (MACI) model of Russell (1999). The model imposes a log-linear vector autoregressive dynamics on the vector of conditional intensities of the multivariate point process and is thus capable of modeling the clustering of events of the same types (self-excitation), as well as the feedback effect transmitting from the previous cluster of events of one type to the future occurrence of events of a different type (mutual excitation). Another example is the Generalized Hawkes model of Bowsher (2007). Rather than implied from an economic model, the model specifications are often proposed on a pragmatic ground, so that the model predictions would match well with the observed stylized facts of the data. Nevertheless, this approach offers a convenient parametric platform for testing Granger causality.

The second approach was developed and briefly illustrated in Chapter two.

It is more favorable than the first approach in that the test conclusion is independent of parametric assumptions. In section 3.4, we will apply the methodology to the study of bankruptcy contagion.

3.2 VAR(1) model

Before converting bivariate point process data into discrete time series of counts, we need to first fix a sampling frequency. We then form a group k count sequence $\{x_t^k\}$ by counting the number of group k events falling in each sampling interval. The two count sequences associated with groups a and b events constitute a bivariate time series of counts $\mathbf{x}_t = \{x_t^a, x_t^b\}$. Unlike the raw point process data, the two count sequences are synchronized in time as the sampling frequency is chosen to be common across all types of events.

Under the vector autoregression of order one (VAR(1) model), the bivariate time series of counts \mathbf{x}_t follows an autoregressive dynamics below:

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \varepsilon_t$$

where $\mathbf{B} = \begin{pmatrix} \beta^{aa} & \beta^{ab} \\ \beta^{ba} & \beta^{bb} \end{pmatrix}$ and the error vector ε_t is a white noise sequence with mean $(0, 0)'$ and covariance matrix $\Sigma = \begin{pmatrix} \sigma^{aa} & \sigma^{ab} \\ \sigma^{ab} & \sigma^{bb} \end{pmatrix}$.

The autoregressive parameter β^{aa} (β^{bb}) captures the persistence of the count sequence x_t^a (x_t^b). On the other hand, the feedback parameter β^{ab} (β^{ba}) picks up the one-lag Granger causal effect from x_{t-1}^b to x_t^a (from x_{t-1}^a to x_t^b). It is obvious that the feedback parameters do not account for all Granger causal dynamics

exhibited by the data. For instance, the VAR(1) model is unable to account for possible Granger causal dynamics spanning beyond one lag. Besides, the discreteness of the count sequences artificially introduces instantaneous Granger causality, which is reflected by the instantaneous correlation $\rho^{ab} = \sigma^{ab} / \sqrt{\sigma^{aa}\sigma^{bb}}$ of the error. It is thus possible that statistical inference on Granger causality in the VAR model framework is sensitive to the model specification (lag order, link function) and the length of a sampling interval.

3.2.1 Spurious causality

In general, spurious causality may arise when the sampling interval is coarser than the actual feedback dynamics. In a bivariate time series context, we will demonstrate the possibility of *spurious causality*, i.e. the process observed at a lower frequency displays Granger causality while the DGP does not exhibit Granger causality. Suppose the DGP is:

$$\begin{aligned}x_t &= ax_{t-1} + u_t, \\y_t &= by_{t-1} + cx_t + v_t,\end{aligned}\tag{3.1}$$

where u_t and v_t are two independent sequences of iid noises. Assume that a and c are not zeros, and that $a + b \neq 0$. Under this model, the process x_t does not Granger cause y_t , although there is contemporaneous correlation between x_t and y_t . In matrix form, the system becomes

$$\begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}.$$

Let $\mathbf{z}_t = (x_t, y_t)'$ and $\varepsilon_t = (u_t, cu_t + v_t)'$. We can then express the bivariate system in reduced form,

$$(\mathbf{I} - \Phi L) \mathbf{z}_t = \varepsilon_t. \quad (3.2)$$

Here, $\Phi = \begin{pmatrix} a & 0 \\ ac & b \end{pmatrix}$ and L is the lag operator (so that $L\mathbf{z}_t = \mathbf{z}_{t-1}$).

Now, let us introduce the temporally aggregated series $\mathbf{z}_t^* = \mathbf{z}_t + \mathbf{z}_{t-1}$ and $\varepsilon_t^* = \varepsilon_t + \varepsilon_{t-1}$. Then, multiply $T(L) := (\mathbf{I} + L)(\mathbf{I} + \Phi L)$ to both sides of (3.2), we obtain

$$(\mathbf{I} - \Phi^2 L^2) \mathbf{z}_t^* = (\mathbf{I} + \Phi L) \varepsilon_t^*.$$

At the lower sampling frequency, it can be shown (Marcellino, 1999) that the right hand side has a moving average representation of order 1. Specifically, there exists an innovation \mathbf{e}_t such that

$$(\mathbf{I} - \Phi^2 L^2) \mathbf{z}_t^* = (\mathbf{I} - \Lambda^2 L^2) \mathbf{e}_t. \quad (3.3)$$

There are two consequences of temporal aggregation: first, the VAR(1) model specification in (3.2) is no longer adequate for the temporally aggregated series \mathbf{z}_t^* , which now follows a VARMA(1,1) structure; second, the autoregressive coefficient matrix is changed from Φ to Φ^2 . The latter change is the source of spurious Granger causality. Since

$$\Phi^2 = \begin{pmatrix} a^2 & 0 \\ ac(a+b) & b^2 \end{pmatrix},$$

we note that $\beta^{(2)} := ac(a+b) \neq 0$ in general (if $a+b \neq 0$), in which case the aggregated model (3.3) indicates feedback from x_{t-2}^* to y_t^* , even though the DGP (3.1) does not suggest any causality from past values of x_t to y_t . From the expression of $\beta^{(2)}$, we see that the autocorrelation of x_t (a), the contemporaneous

cross-correlation of x_t and y_t (c) and the way the autocorrelations of x_t and y_t interact with one another ($a + b$) all contribute to spurious Granger causality.

3.2.2 Spurious non-causality

Spurious non-causality may also arise when the sampling interval is coarser than the actual feedback dynamics. *Spurious non-causality* occurs when the process observed at a lower frequency displays no Granger causality while the DGP exhibits Granger causality. Suppose the DGP is instead:

$$\begin{aligned}x_t &= ax_{t-1} + u_t, \\y_t &= by_{t-1} + dx_{t-1} + v_t.\end{aligned}\tag{3.4}$$

Let $\mathbf{z}_t = (x_t, y_t)'$, $\varepsilon_t = (u_t, v_t)'$ and $\Phi = \begin{pmatrix} a & 0 \\ d & b \end{pmatrix}$, with $a, b, d \neq 0$. The DGP (3.4) can then be expressed as (3.2). The temporally aggregated series is thus given by (3.3), with

$$\Phi^2 = \begin{pmatrix} a^2 & 0 \\ d(a+b) & b^2 \end{pmatrix}.$$

It is possible that $\gamma^{(2)} := d(a+b) = 0$. In other words, spurious Granger non-causality occurs in the temporally aggregated model for \mathbf{z}_t^* when the autoregressive coefficients of x_t and y_t cancel one another ($a+b=0$).

3.3 Bivariate ACI model

Point process models offer a better alternatives that respect the raw event times. Suppose we are given with the point process data (t_i, y_i) . The specification of a

bivariate ACI model with constant baseline intensity function is given below. For $k = a, b$, the conditional intensity function given as

$$\lambda^k(t) = \exp(\omega^k + \phi_{N(t^-)}^k).$$

The process $\phi_i = (\phi_i^a, \phi_i^b)'$ is assumed to follow an autoregressive structure which is updated at each event time. Specifically, for a bivariate ACI(1,1) model, we have

$$\phi_i = \sum_{k=a,b} \mathbf{a}^k \varepsilon_i^k 1(y_i = k) + \mathbf{B} \phi_{i-1}$$

where $\mathbf{a}^k = (\alpha^{ka}, \alpha^{kb})'$, $\mathbf{B} = (\beta^{jj'})_{j,j' \in \{a,b\}}$ and

$$\varepsilon_i^k = 1 - \int_{U_i^k} \lambda^k(t) dt.$$

The second term of the error is obtained by integrating the conditional intensity for group k bankruptcies over the most recent group k bankruptcy time duration $U_i^k = [t_{N^k(t_i)-1}^k, t_{N^k(t_i)}^k]$.

The MACI(1,1) model offers one of the many possible parametric model specifications on the conditional intensities of a multivariate point process. Although there exists considerable generalization to the model specifications (e.g. more lags, more general baseline intensity functional form), the family of MACI models is still quite restrictive in the sense that a log-linear autoregressive dependence structure (with exponentially decaying auto- and cross-correlations) is imposed on the conditional intensities. Such model specification may not fit well to realized point process data. It is difficult to find an MACI model specification that adequately fits (in terms of goodness-of-fit test results) the observed high-frequency tick data (Russell, 1999).

3.4 Nonparametric Granger causality test

The nonparametric Granger causality test does not depend on a strong model specification and any particular sampling frequency. In its most general form, the test is able to detect all kinds of pairwise dependence between the residual processes of the two marginal simple point processes. There is a simplified and more powerful Q test that is sensitive to persistent pairwise dependence between the residual processes. The latter test is more useful and powerful because credit events are usually observed to cluster together over time. For instance, during economic recessions, the default/failure of a company is often followed by defaults/failures of other companies in a short time period. Such credit contagion may be confined to a particular industry that bears the brunt of a recession, but it may well spread from one sector to another. The Q test works by summarizing all cross correlations of the residual components over all possible lags, meanwhile controlling for the autocorrelations of each marginal point process. It can be shown that, for bivariate simple point process, the existence of pairwise Granger causality is equivalent to non-zero cross correlations of the residual processes.

3.5 Empirical results

We want to show that, during economic recessions, credit contagion spreads from the bottom to the top of a typical supply chain. When determining the periods of economic recession, we refer to the National Bureau of Economic Research (NBER)'s published list of U.S. Business Cycle Expansions and Contractions. According to NBER's classification, our bankruptcy data contain three

periods of recession: July 1990 to March 1991, March 2001 to November 2001, and December 2007 to June 2009. The NBER's definition of economic recession depends on macroeconomic indicators such as real GDP and unemployment rate², but we believe that chapter 11 filings usually occur with a time delay - the impact of recession may only be felt by companies after some time and eventually result in bankruptcies of invulnerable ones with a time lag. We therefore carry out our empirical analysis over sampling windows that are expanded to the right by different extents

Limitations: We only focus on the analysis of timing of Chapter 11 filing cases for large firm. I use the "bankruptcies of U.S. firms, 1980–2010" dataset to study credit contagion. The dataset is maintained by Professor Lynn LoPucki of UCLA School of Law. The dataset records, among other entries, the filing dates of Chapter 11 and the Standard Industrial Classification (SIC) codes of big bankrupting firms. In this analysis, a credit event is defined as the occurrence of bankruptcy event(s). Other events (default, firm exit, merger and acquisition) are not considered in this empirical study.

3.5.1 Bankruptcy Contagion: upstream vs downstream

To test for bankruptcy contagion along a supply chain, we make use of the SIC division codes to classify the bankrupting firms into two groups. The ordering of SIC division codes naturally reflects the hierarchy of a typical supply chain. Hence, we divide all bankrupting firms into two groups according to their SIC codes: the first group (upstream) consists of five groups of industries: agricul-

²A recession is defined by NBER as "a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales".

tural (A), mining (B), construction (C), manufacturing (D), and transportation, communications, electric, gas (E); while the second group (downstream) consists of four groups of industries: wholesale (F), retail (G), finance, insurance and real estate (H), and services (I).

The results in Tables B.3 to B.8 show conflicting Granger causal directions for VAR(1) models³ with different sampling frequencies (Tables B.3 to B.5 for recessions and crises; Tables B.6 to B.8 for recoveries and expansions). Due to the continuous time feature, the ACI(1,1) model is powerful in picking up Granger causality, although the estimates suggest that the DGP exhibits non-stationarity (see Table B.1 for recessions and crises, and Table B.2 for non-recession periods). Figures B.1 and B.2 display the results of the nonparametric Q^s test which are in agreement with those of the ACI(1,1) models. The results show that there is significant Granger causality from the bottom to the top of a supply chain, which indicates that bankruptcy contagion is demand-driven.

3.5.2 Bankruptcy Contagion: Wall Street vs Main Street

It is well believed that the financial industry affects other sectors of the economy during the subprime mortgage crisis. To check the proposition that bankruptcy contagion spreads from financial to non-financial sectors, we divide the bankrupting firms into two groups according as the SIC division is H (Wall Street) and not H (Main Street). We conducted the nonparametric Granger causality test over both the recession and expansionary/non-recession periods. The 1990 recession only contains four bankruptcy filings from the financial sec-

³Vector autoregressions with higher lag orders were also estimated (results are available upon request). They still give conflicting results.

tor, so the test is not reliable. According to Figure B.3, we found marginally significant (at 5%, for bandwidths B between 90 and 120 days) Granger causality from Wall Street bankruptcies to Main Street bankruptcies during the subprime mortgage crisis. We also run the test over the 2001 recession (March 2001 to November 2002, extended), and found no evidence of a Wall-Street-to-Main-Street bankruptcy contagion in the control sample periods. However, we documented very significant Wall-Street-to-Main-Street bankruptcy contagion during the brief period after the subprime mortgage crisis (July 2009 to November 2011) over all bandwidths B , as well as during the recovery period from December 2001 to November 2007 over shorter time spans (at 5%, for $B \leq 90$ days), according to Figure B.4. This is possibly due to the slight delay in bankruptcy occurrences relative to the business cycle.

CHAPTER 4
DIAGNOSTIC CHECKS FOR MULTIVARIATE PARAMETRIC
INTENSITY MODELS

4.1 Introduction

Various events occur in the economy stochastically. No matter it is a single transaction of stocks in a stock exchange, the onset of unemployment of an individual, the change of interest rate, or the bankruptcy of a corporation, the occurrences of events are, more often than not, uncertain. With the hope of gaining a better understanding of how and why events occur, economists devise counting process models for datasets containing irregularly spaced event times that are often associated with other attributes at those instances. A model that is adequate for the data can potentially provide an explanation to the patterns of event occurrences observed in the past, and even shed light on possible trends in the future.

Counting process models have a deep root in biostatistics (see Andersen et al, 1993 for an introduction). The datasets are often a panel of subjects under study with different start and end times (sometimes truncated or censored) of durations of various lengths. Statisticians find it most natural to specify a counting process model by associating each subject with a conditional intensity function, which gives the probability of events happening in the next instance conditional on past history. It takes the more well-known name of hazard function if the counting process is a renewal process. Proportional hazard models and accelerated failure time models are two important classes of models in this category. Estimation and validation of parametric models on panel data are well de-

veloped in biostatistics and econometrics. Applications in economics and finance are abundant too - some examples include: duration models of unemployment spells in labor economics (Lancaster, 1979; see Kiefer, 1988 for a survey) and reduced-form intensity models of defaults in credit risk analysis (Jarrow and Turnbull, 1995; see Lando, 2004, Chapter 5 for an introduction).

Sometimes, instead of taking a panel approach, it is more convenient to look at the data from a time series perspective. Events of the same type may occur recurrently over time and it is far more parsimonious to have a single conditional intensity function that governs the evolution of multiple homogeneous events over time. These recurrent event counting process models can also be found in economics and finance. The autoregressive conditional duration (ACD) model (Engle and Russell, 1998) and its variants are particularly useful in capturing the stylized fact of duration clustering usually exhibited in high-frequency trade and quote datasets. From a modeler's point of view, they are appealing because the dynamic structure is directly imposed on conditional durations, and interarrival event durations are observable from the data. The autoregressive conditional intensity (ACI) model (Russell, 1999) takes an alternative approach by modeling the conditional intensity of events. Although less concrete than conditional duration as it seems, conditional intensity can be a more convenient building block in multivariate extensions of a counting process model. Another example is the generalized Hawkes model (Bowsher, 2007) which finds its use in multivariate trade and quote arrivals. In credit risk literature, Giesecke (2008) proposed a top-down approach to model the defaults of a portfolio of firms. Parametric models are often estimated by practitioners by typical techniques such as maximum likelihood estimation (MLE) and evaluated by standard time series diagnostics tools such as Box-Pierce or Ljung-Box tests applied on gener-

alized residual series out of an estimated intensity model. This chapter focuses on the model validation of intensity models for recurrent events.

Empirical researchers in econometrics and finance have been relying on handy time series diagnostic tests to check for model adequacy of intensity models. Engle and Russell (1998) tested for serial correlations of the residual sequence by Ljung-Box test which, together with the excess dispersion test, indicates whether an ACD model is adequate for the data. To capture the stylized fact of default clustering, Das, Duffie, Kapadia and Saita (2007) estimated a Cox model and conducted statistical tests on the time transformed counting process, which should be a standard Poisson process if the intensity model is adequate for the default data. One of the tests is to test for serial correlations by fitting an AR model to the sequence of numbers of defaults in equal-sized windows after time transformation. The properties of the popular portmanteau tests are of course well known for standard time series models with fixed and discrete time intervals such as vector ARMA and GARCH models and their variants. However, there are not many theoretical results on the asymptotic distribution of these test statistics, nor any empirical performance analysis of these tests, applied on a general multivariate parametric continuous-time intensity model with parameter uncertainty. (Kwok and Li, 2008 discussed these topics but was mainly confined to univariate autoregressive conditional intensity models only.) This chapter aims at filling this gap in the literature, by deriving a large-sample distribution theory for generalized residual autocorrelations and proposing statistical test procedures that check for model adequacy. The test procedures thus derived are theoretically justified for a wide class of multivariate parametric recurrent-event intensity models.

4.2 Parametric multivariate point process models

We are given a multivariate point process N with the following details:

Event time, event type pairs: $\{(T_1, M_1), (T_2, M_2), \dots\}$. Set $t_0 = 0$.

Event types: $1, \dots, K$. E.g. $M_i = 1$ if the i^{th} event is of type 1.

Pooled counting process: $N(t) = \sum_{i=1}^{\infty} 1(t_i \leq t)$

Marginal counting processes: $N^k(t) = \sum_{i=1}^{\infty} 1(t_i \leq t) 1(k_i = k), k \in \{1, \dots, K\}$

Natural filtration: $\mathcal{F} = (\mathcal{F}_t)$ with $\mathcal{F}_t = \sigma\{(t_1, k_1), (t_2, k_2), \dots, (t_{N(t)}, k_{N(t)})\}, t \geq 0$.

Assume that \mathcal{F}_t is right continuous.

Probability space: $(\Omega, \mathcal{F}, \mathbb{P})$

Conditional intensity function: $\lambda(t|\mathcal{F}_{t-}) = \left(\lambda^1(t|\mathcal{F}_{t-}), \lambda^2(t|\mathcal{F}_{t-}), \dots, \lambda^K(t|\mathcal{F}_{t-})\right)'$ ¹

Time horizon with observations: $[0, T]$

Let $n = N(T)$. We assume that $n \rightarrow \infty$ as $T \rightarrow \infty$.

Assume that N is orderly and integrable, so that there is at most one event occurring at any instant almost surely, and that Γ does not explode, so that $\mathbb{E}[N(t)] < \infty$ for any finite t .

Suppose that Γ is generated by a parametric model which can be specified in

¹A conditional (or stochastic) intensity process $\lambda(t|\mathcal{F}_{t-})$ is a nonnegative, \mathcal{F}_t -progressive process such that for all $t \geq 0$, $\int_0^t \lambda(r|\mathcal{F}_{r-}) dr < \infty$ and for all \mathcal{F}_t -predictable processes $C(t)$, $E\left[\int_0^\infty C(t) dN(t)\right] = E\left[\int_0^\infty C(t) \lambda(t|\mathcal{F}_{t-}) dt\right]$ holds.

terms of its conditional intensity. For some $\theta = \theta_0$ in the parameter space $\Theta \in R^D$,

$$\begin{aligned}\lambda(t|\mathcal{F}_{t-}) &= \lambda(t|\mathcal{F}_{t-}; \theta_0) \\ &= \left(\lambda^1(t|\mathcal{F}_{t-}; \theta_0), \lambda^2(t|\mathcal{F}_{t-}; \theta_0), \dots, \lambda^K(t|\mathcal{F}_{t-}; \theta_0) \right)'. \end{aligned} \quad (4.1)$$

4.3 Random Time Change and Characterization of Multivariate

Intensity Models

For a general marked point process the joint distribution of (t, k) given \mathcal{F}_t can be characterized by the following composite function:

$$\lambda^*(t, k|\mathcal{F}_{t-}) = h(t|\mathcal{F}_{t-})f(k|\mathcal{F}_{t-}, t),$$

where $h(t|\mathcal{F}_{t-})$ is the intensity function of the pooled process and $f(k|\mathcal{F}_{t-}, t)$ is the density function of the mark given the event location and the process history (Daley, Vere-Jones, 2002, p.249). For a multivariate point process with finite mark space $\{1, \dots, K\}$,

$$h(t|\mathcal{F}_{t-}) = \sum_{j=1}^K \lambda^j(t|\mathcal{F}_{t-})$$

and

$$f(k|\mathcal{F}_{t-}, t) = \frac{\lambda^k(t|\mathcal{F}_{t-})}{\sum_{j=1}^K \lambda^j(t|\mathcal{F}_{t-})}.$$

Here, the conditional density function $f(k|\mathcal{F}_{t-}, t)$ should be understood as a conditional probability mass function. Therefore, for a multivariate point process model, there are two aspects of model adequacy which a comprehensive diagnostic procedure should check for:

1. The intensity function of the pooled process should model the timing of all events (regardless of event type) adequately, i.e. $h(t|\mathcal{F}_t^-; \theta) = h(t|\mathcal{F}_t^-; \theta_0)$ for some $\theta \in \Theta$; and
2. Given that an event occurs at time t , the conditional distribution of event types should model the observed event types adequately, i.e. $f(k|\mathcal{F}_t^-, t; \theta) = f(k|\mathcal{F}_t^-, t; \theta_0)$ for some $\theta \in \Theta$.

Let us recall the random time change theorem which is the basis for statistical tests on intensity models.

Theorem 15 (Meyer, 1971) *Suppose that the K sequences of arrival times $\{T_i^1\}_{i=1}^\infty, \{T_i^2\}_{i=1}^\infty, \dots, \{T_i^K\}_{i=1}^\infty$ (where T_i^j is the time when the i^{th} event of type j occurs) are generated by a multivariate point process $\Gamma = \{N^1(t), N^2(t), \dots, N^K(t)\}_{t \geq 0}$ with absolutely continuous compensator $\Lambda = \{\Lambda_1(t), \Lambda_2(t), \dots, \Lambda_K(t)\}_{t \geq 0}$, such that $\Lambda_j(t) \rightarrow \infty$ as $t \rightarrow \infty$ for all $j = 1, 2, \dots, K$. Then, the time transformed process $\tilde{\Gamma} = \{N^1(\Lambda_1^{-1}(t)), N^2(\Lambda_2^{-1}(t)), \dots, N^K(\Lambda_K^{-1}(t))\}_{t \geq 0}$ is a multivariate standard Poisson process, i.e. each $N^j(\Lambda_j^{-1}(\cdot))$ is a univariate Poisson process with rate 1, and the K component processes $N^j(\Lambda_j^{-1}(\cdot))$ are independent over event types $j = 1, 2, \dots, K$.*

As an easy consequence of the above theorem, the time durations between consecutive events of the same type from the time transformed process $\tilde{\Gamma}$ are iid exponential distributed with rate 1. This is stated formally in the corollary below:

Corollary 16 *Suppose that the absolutely continuous marked compensator Λ_j of the point process defined in the above theorem admits a predictable intensity process $\lambda^j(t)$*

as its density. Then, the time duration between the $(i - 1)^{th}$ and i^{th} events of type j

$$x_i^j = \tilde{T}_i^j - \tilde{T}_{i-1}^j = \Lambda_j(T_i^j) - \Lambda_j(T_{i-1}^j) = \int_{T_{i-1}^j}^{T_i^j} \lambda^j(t) dt$$

follows iid exponential(1) distribution. Note that x_i^j are independent over all i and j .

In the context of parametric intensity models (4.1), by defining $\varepsilon_i^j = 1 - x_i^j$, we then get a sequence $\{\varepsilon_i^j : i = 1, 2, \dots; j = 1, \dots, K\}$ of innovations that drives the dynamics of the intensity model. From corollary 16, we learn that ε_i^j ($i = 1, 2, \dots; j = 1, \dots, K$) are iid with mean 0 and a distribution that is a reflected and shifted exponential(1). This provides the basis of goodness-of-fit diagnostics.

The following theorem gives an equivalent characterization of a multivariate standard Poisson process. By focusing on the pooled point process, the alternative perspective the theorem provides simplifies the goodness-of-fit problem of a multivariate intensity model to that of a univariate one with marks. First, we need a lemma that reiterates the memoryless property of Poisson process.

Lemma 17 *Let U_i ($i = 1, 2, \dots$) be the i^{th} event time of a univariate standard Poisson process. Suppose we know that $U_i \leq c < U_{i+1}$ for a fixed time point c . Then, the distribution of $(U_{i+1} - c)$ is exponential(1).*

Proof. Given that $U_i \leq c < U_{i+1}$, there are no events in $[c, U_{i+1})$, and so by stationarity of Poisson process, $P(U_{i+1} - c > t | U_i \leq c < U_{i+1}) = P(U_{i+1} > c + t | U_{i+1} > c) = P(U_1 > t) = e^{-t}$. ■

Theorem 18 *The following three statements are equivalent:*

1. The process Φ with type j event times $\{U_i^j\}_{i=1}^\infty$ and time origin $U_0^j = 0$ ($j = 1, 2, \dots, K$) is a multivariate standard Poisson process of dimension K ;

2. The time duration $Y_i^j = U_i^j - U_{i-1}^j$ ($i = 1, 2, \dots; j = 1, 2, \dots, K$) between the $(i-1)^{\text{th}}$ and i^{th} events of type j follows iid exponential(1) distribution;
3. Let $Y_i = U_i - U_{i-1}$ ($i = 1, 2, \dots$) be the i^{th} duration of the pooled process $\tilde{\Phi} = \{N(t)\} = \{\sum_{j=1}^K N^j(t)\}$ of Φ and W_i be the mark of the i^{th} event. Then,
 - (a) Y_i follows iid exponential(K) distribution;
 - (b) W_i follows iid discrete uniform distribution, taking values $1, 2, \dots, K$ with probability $\frac{1}{K}$ each; and
 - (c) Y_i and W_i are independent for all $i = 1, 2, \dots$

Proof. (1) \iff (2) is standard. It follows from the independent and stationary increment properties of a Poisson process.

(2) \Rightarrow (3): To show (3a), it suffices to show that, for all $j = 1, 2, \dots, K$, the conditional distribution of $U_{N^j(U_i)+1}^j - U_i$ given U_i follows exponential(1), since then the unconditional distribution of $U_{N^j(U_i)+1}^j - U_i$ follows exponential(1) too and hence $Y_{i+1} = \min\{t - U_i : N(t) > N(U_i)\} = \min\{U_{N^j(U_i)+1}^j - U_i : j = 1, 2, \dots, K\}$ follows exponential(K). Without loss of generality, suppose $W_i = 1$. If $W_{i+1} = W_{i-1} = 1$, then $(U_{N^j(U_i)+1}^j - U_i)|U_i = Y_{N^1(U_i)+1}^1|U_i = Y_{N^1(U_i)+1}^1 \sim \exp(1)$. If $W_{i+1} = j \neq 1$, then by the memoryless property of Poisson process (see the above lemma), $(U_{N^j(U_i)+1}^j - U_i)|U_i \sim \exp(1)$ as well. (3b) follows from symmetry: given Y_i , the probability that $W_i = j$ is $\frac{\exp(-Y_i K)}{K \exp(-Y_i K)} = \frac{1}{K}$ for all $j = 1, 2, \dots, K$, and (3c) follows from the fact that the conditional distribution of W_i given Y_i is independent of Y_i as seen from the proof of (3b).

(3) \Rightarrow (2): Without loss of generality, suppose $W_i = 1$. Then, given $(U_i, W_i = 1)$,

$$Y_{N^1(U_i)}^1 = \begin{cases} U_{i+1} & \text{with prob. } \frac{1}{K} \\ U_{i+1} + U_{i+2} & \text{with prob. } \frac{K-1}{K} \frac{1}{K} \\ U_{i+1} + U_{i+2} + U_{i+3} & \text{with prob. } \left(\frac{K-1}{K}\right)^2 \frac{1}{K} \\ \vdots & \vdots \end{cases}$$

By (3a), $U_{i+1} + U_{i+2} + \dots + U_{i+r}$ follows a gamma distribution with parameters r (shape) and K (scale). The moment generating function of $Y_{N^1(U_i)}^1$ given $(U_i, W_i = 1)$ is:

$$\begin{aligned} & \mathbb{E}\left(\exp(tY_{N^1(U_i)}^1) \middle| U_i, W_i = 1\right) \\ &= \frac{K}{K-t} \frac{1}{K} + \left(\frac{K}{K-t}\right)^2 \frac{K-1}{K} \frac{1}{K} + \left(\frac{K}{K-t}\right)^3 \left(\frac{K-1}{K}\right)^2 \frac{1}{K} + \dots \\ &= \frac{\frac{K}{K-t} \frac{1}{K}}{1 - \frac{K}{K-t} \frac{K-1}{K}} \\ &= \frac{1}{1-t}. \end{aligned}$$

Similarly, $Y_{N^j(U_i)}^j$ given $(U_i, W_i = j)$ follows iid exponential(1) for all $j = 1, 2, \dots, K$ and all U_i . Since the conditional distribution is independent of U_i and W_i , the unconditional distribution of Y_k^j also follows iid exponential(1). ■

Because of the above characterization theorem, it is sufficient to base the goodness-of-fit problem of model (1) on checking the following three aspects:

1. The durations x_i between the $(i-1)^{th}$ and i^{th} events of the *pooled* time transformed point process, defined by

$$x_i = \tilde{T}_i - \tilde{T}_{i-1} = \int_{\tilde{T}_{i-1}}^{\tilde{T}_i} \sum_{j=1}^K \lambda^j(t) dt,$$

are iid exponential(K).

2. The marks m_i of the *pooled* time transformed point process are iid discrete uniform over $1, 2, \dots, K$.
3. x_i and m_i are independent for all $i = 1, 2, \dots$

One may instead approach the goodness-of-fit problem by examining the iid property of the marked durations x_i^j from an estimated model after time transformation. However, there are at least two advantages of our three-step procedure over this approach. First, unlike traditional time series analysis, the K sequences of marked durations x_i^j from a multivariate point process are asynchronously observed by nature. For instance, there is no straightforward way, if there is any, of defining a lag-one cross correlation of the two sequences $\{x_i^1\}$ and $\{x_i^2\}$, because the realized i^{th} marked duration between events of type 1 may be observed before or after, and can be very far away in time from, that between events of type 2. The total numbers of realized x_i^1 and x_i^2 can be very different too, which may introduce problems that compromise the asymptotic property of the corresponding test statistic that is a function of the realized marked durations. The three-step approach does not have this problem. Second, it is not clear what alternative hypothesis a test statistic based on realized marked durations is testing against. At best, a tailor-made test statistic has to be considered if, for instance, the alternative hypothesis is that occurrence of event 1 excites the occurrence of event 3, but this can be more systematically tested against by a multiple runs test (step 2) on the sequence of marks \tilde{M}_i under the three-step approach without the need for a tailor-made test statistic. Moreover, against the special alternative hypothesis of a two-state dependent first order stationary Markov chain, runs test can be shown to be the most powerful unbiased test for randomness of an ordinal-valued sequence (Lehmann and Romano, 1971, p.146).

4.3.1 Generalized Residual Autocorrelations Test

The purpose of step one is to check that the time transformed durations from the pooled process of an estimated intensity model has the iid exponential(1) properties. Now, let us concentrate on testing for serial correlations. To this end, we need to derive the asymptotic distributions of a vector of autocorrelations of generalized residuals from an estimated intensity model.

By Meyer's (1971) random time change theorem,

$$x_i(\theta_0) = \int_{T_{i-1}}^{T_i} \sum_{j=1}^K \lambda^j(t|\mathcal{F}_{t-}; \theta_0) dt,$$

$i = 1, \dots, n$ follow iid exponential distribution with rate K . After replacing all x_i by $x_i(\theta_0)$, we can apply the above tests for a Poisson process as tools of checking model adequacy of a general parametric multivariate point process, provided that the true value of parameter θ_0 is known.

In reality, however, θ_0 is unknown and has to be estimated by, say, maximum likelihood estimation, from the data. Let $\hat{\theta}$ be an estimator of θ_0 . Because of parameter uncertainty, $x_i(\hat{\theta}), i = 1, \dots, n$ are no longer iid standard exponential, and the asymptotic distributions of all the test statistics considered above are no longer the same. For notational simplicity, let $\varepsilon_i = 1 - Kx_i(\theta_0)$ and $\hat{\varepsilon}_i = 1 - Kx_i(\hat{\theta})$. The sample mean is $\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i$.

Let us impose the following assumptions on the parametric model $\lambda(t|\mathcal{F}_{t-}; \theta)$:

Assumption 1: The pooled point process is stationary and ergodic with absolutely continuous compensator.

Assumption 2: $\hat{\theta}$ is the maximum likelihood estimator of θ_0 that maximizes

the log-likelihood function

$$\ell(\theta) = \sum_{j=1}^K \left[- \int_0^T \lambda^j(t|\mathcal{F}_{t^-}; \theta) dt + \sum_{i=1}^n \log \lambda^j(t_i|\mathcal{F}_{t^-}; \theta) \right]. \quad (4.2)$$

Moreover, the regularity conditions in Ogata (1978) that guarantee the existence and uniqueness of $\hat{\theta}$ are assumed on the pooled process with intensity $h(t; \theta) = \sum_{j=1}^K \lambda^j(t|\mathcal{F}_t; \theta)$.

Assumption 3: The fourth moment $\mathbb{E}(\varepsilon_i^4)$ exists and is finite.

Let us consider the serial autocorrelations test with parameter uncertainty. We are interested in testing

$$H_0 : \hat{\varepsilon}_i, i = 1, \dots, n, \text{ are serially uncorrelated} \quad \text{vs}$$

$$H_1 : \hat{\varepsilon}_i, i = 1, \dots, n, \text{ are serially correlated.}$$

To this end, we construct a portmanteau test statistic which is composed of a sequence of lag m sample autocorrelations of $\hat{\varepsilon}_i$ for $m = 1, \dots, M$, defined by

$$\hat{\rho}_m = \frac{\sum_{i=1}^{n-m} (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})(\hat{\varepsilon}_{i+m} - \bar{\hat{\varepsilon}})}{\sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2}.$$

Under the null hypothesis, the sample mean $\bar{\hat{\varepsilon}}$ converges in probability to $E(\varepsilon_i) = 0$ and the sample variance $\sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2/n$ converges in probability to $\text{Var}(\varepsilon_i) = 1$ as $n \rightarrow \infty$. Hence, $\hat{\rho}_m - \hat{r}_m \xrightarrow{P} 0$ as $n \rightarrow \infty$, where

$$\hat{r}_m = \frac{\sum_{i=1}^{n-m} \hat{\varepsilon}_i \hat{\varepsilon}_{i+m}}{n}.$$

Therefore, it suffices to investigate the joint distribution of the sample autocovariances \hat{r}_m for $m = 1, \dots, M$. Define $\hat{\mathbf{r}} = (\hat{r}_1, \hat{r}_2, \dots, \hat{r}_M)'$ and its population counterpart $\mathbf{r} = (r_1, r_2, \dots, r_M)'$, where $r_m = \sum_{i=1}^{n-m} \varepsilon_i \varepsilon_{i+m}/n$. Then, a Taylor expansion around the true parameter θ_0 yields:

$$\hat{\mathbf{r}} = \mathbf{r} + \frac{\partial \mathbf{r}}{\partial \theta} (\hat{\theta} - \theta_0) + O_p(n^{-1}).$$

Denote the probability limit of $-\partial \mathbf{r} / \partial \theta$ by \mathbf{X} , i.e. $\mathbf{X} = -\mathbb{E} [\partial \mathbf{r} / \partial \theta]$. Then, as $n \rightarrow \infty$, we have

$$\hat{\mathbf{r}} \xrightarrow{P} \mathbf{r} - \mathbf{X}(\hat{\theta} - \theta_0).$$

First, note that $\sqrt{n} \mathbf{r}$ converges in distribution to $N(0, \mathbf{I}_M)$, where \mathbf{I}_M is the $(M \times M)$ identity matrix (Hannan, 1967). Second, by assumption 2 and the property of maximum likelihood estimator for point process models (Ogata, 1978), $\sqrt{n}(\hat{\theta} - \theta_0)$ converges in distribution to $N(0, \mathbf{G}^{-1})$ as $n \rightarrow \infty$, where $\mathbf{G} = -\mathbb{E} [n^{-1} \partial^2 \ell / \partial \theta \partial \theta']$ is the information matrix. By the martingale central limit theorem, $\sqrt{n} \hat{\mathbf{r}}$ has an asymptotic normal distribution. The result is stated in the following theorem.

Theorem 19 *Under Assumptions 1-3, $\sqrt{n} \hat{\mathbf{r}} \xrightarrow{d} N(0, \mathbf{V})$ where*

$$\mathbf{V} = \mathbf{I}_M - \mathbf{X} \mathbf{G}^{-1} \mathbf{X}', \quad (4.3)$$

$\mathbf{X} = -\mathbb{E} [\partial \mathbf{r} / \partial \theta]$ is an $(M \times D)$ matrix with the (m, d) entry given by

$$x_{md} = -\frac{1}{n} \sum_{i=m+1}^n \mathbb{E} \left[\varepsilon_{i-m} \frac{\partial}{\partial \theta_d} \log h(T_i; \theta_0) \right], \quad (4.4)$$

and $\mathbf{G} = -\mathbb{E} [n^{-1} \partial^2 \ell(\theta_0) / \partial \theta \partial \theta']$ is a $(D \times D)$ invertible matrix with the (p, q) entry given by

$$g_{pq} = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial}{\partial \theta_p} \log h(T_i; \theta) \frac{\partial}{\partial \theta_q} \log h(T_i; \theta) \right]. \quad (4.5)$$

Corollary 20 *The portmanteau test statistic $Q_1(M) = n \hat{\rho}' \mathbf{V}^{-1} \hat{\rho}$ converges to a chi-squared distribution with M degrees of freedom.*

To implement the tests, it is necessary to estimate the asymptotic variance \mathbf{V} from the data. The following theorem provides a consistent estimator for \mathbf{V} .

Theorem 21 Define $\hat{\mathbf{V}} = \mathbf{I}_M - \hat{\mathbf{X}}\hat{\mathbf{G}}^{-1}\hat{\mathbf{X}}'$, where $\hat{\mathbf{X}}$ is an $(M \times D)$ matrix with the (m, d) entry given by

$$\hat{x}_{md} = -\frac{1}{n} \sum_{i=m+1}^n \hat{\varepsilon}_{i-m} \frac{\partial}{\partial \theta_d} \log h(t_i; \hat{\theta}),$$

and $\hat{\mathbf{G}}$ is a $(D \times D)$ matrix with the (p, q) entry given by

$$\hat{g}_{pq} = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta_p} \log h(t_i; \hat{\theta}) \frac{\partial}{\partial \theta_q} \log h(t_i; \hat{\theta}).$$

Then, $\hat{\mathbf{V}} \xrightarrow{P} \mathbf{V}$.

A couple of remarks on the estimation of the asymptotic covariance matrix \mathbf{V} are in order. First, note that $\hat{\mathbf{G}}$ is readily available from maximization routines as it is also the hessian matrix of the log-likelihood function evaluated at $\hat{\theta}$. To obtain $\hat{\mathbf{X}}$, we need the derivatives of the log conditional intensity function evaluated at t_i and $\hat{\theta}$. For Cox models, they are simply equal to the corresponding covariates. Another important particular case is where the log conditional intensity function admits an ARMA representation as in the univariate ACI(p, q) models. Then, it can be shown along the same arguments as in Box and Pierce (1970) that, by choosing $M = O(\sqrt{n})$, $\mathbf{G} = \mathbf{X}'\mathbf{X}$ asymptotically and hence \mathbf{V} is idempotent with $p + q$ degrees of singularity in the limit. The asymptotic idempotence property saves us from evaluating \mathbf{V} and implies that the distribution of $n\hat{\rho}'\hat{\rho}$ converges to a chi-square distribution with $M - p - q$ degrees of freedom. A third possibility is that the conditional intensity function is modeled by a Markov process such as jump diffusion, then by Markov property it is clear from (C.1) that $\mathbf{X} = \mathbf{0}$, so that $\mathbf{V} = \mathbf{I}_M$ and hence the portmanteau test statistics are distributed as if there is no parameter uncertainty.

4.4 Simulations

To examine the performance of the autocorrelations test proposed in the last section, we conduct simulations to get empirical sizes of the test under different situations. The models of interest are as follows.

Cox model:

$$h(t; \theta) = \exp\{a_0 + a_1 X(t) + a_2 Y(t)\}$$

where $\theta = (a_0, a_1, a_2)$; $X(t)$ and $Y(t)$ are independent, exogenous AR processes: for all integers i :

$$X(i) = bX(i-1) + V(i)$$

$$Y(i) = cY(i-1) + W(i)$$

with iid normal innovations $V(i)$ and $W(i)$.

Univariate ACI(1, 1) model:

$$h(t; \theta) = \exp\{\omega + \phi_{i-1}\}, \text{ for } t \in (t_{i-1}, t_i]$$

$$\phi_i = \alpha\phi_{i-1} + \beta\varepsilon_i$$

$$\varepsilon_i = 1 - \int_{t_{i-1}}^{t_i} h(t; \theta) dt$$

where $\theta = (\omega, \alpha, \beta)$; $\int_{t_{i-1}}^{t_i} h(t; \theta) dt$ are iid exponential(1).

Bivariate ACI(1, 1) model defined in section 3.3.

For the size experiments, we consider the following data generating processes.

DGP1: Univariate ACI(1,1) model with $\theta = (0.5, 0.9, 0.1)$. DGP2: Cox model with $\theta = (0.5, 0.2, 0.3)$ and AR coefficients $b = 0.5, c = 0.6$. DGP3: Bivariate

ACI(1,1) model with $(\omega^a, \omega^b) = (-10, 10)$, $\mathbf{a}^a = \mathbf{a}^b = (0.1, 0.1)'$ and $\mathbf{B} = \begin{pmatrix} 0.5 & 0.1 \\ 0.0 & 0.7 \end{pmatrix}$

We present the empirical size results of the various portmanteau test obtained after 1000 simulations. The nominal size of the tests is set to be 0.05. The empirical sizes of various tests are plotted against the maximum number of lags, M , when the DGP is the ACI(1,1) model and Cox model, respectively. We see that the portmanteau test $Q_1(M)$ considered in this chapter comes close to the nominal level after adjusting the degrees of freedom of its chi-square distribution to $M - 2$ for ACI(1,1) model. Its performance becomes even more promising as the sample length increases to $n = 400$. Similar results hold for Cox model except that the degree of freedom of the chi-square distribution for the $Q_1(M)$ test remains at M . For the bivariate ACI(1,1) model, we conduct Box-Pierce and Ljung-Box tests on the residual sequence from the pooled point process (regardless of event types) as well as the residual sequence of each of the two marginal point processes. The plots show that the empirical rejection rates of the Ljung-Box test are close to the nominal level at M degrees of freedom of the chi-square distribution, while the Box-Pierce test tends to be increasingly undersized as M increases. The satisfactory size performance of the Ljung-Box test suggests that the parameter uncertainty correction of $Q_1(M)$ on the estimated bivariate ACI(1,1) model seems to be minor and not worth the effort.

To study the power of the portmanteau test, we now assume the DGP to be a bivariate ACI(1,1) model with strong two-way feedback, having parameters $(\omega^a, \omega^b) = (-1, 1)$, $\mathbf{a}^a = \mathbf{a}^b = (0.1, 0.1)'$ and $\mathbf{B} = \begin{pmatrix} 0.5 & 0.8 \\ -0.8 & 0.7 \end{pmatrix}$. We then intentionally estimate a univariate ACI(1,1) model to each of the marginal point processes without accounting for the two-way feedback. Both the Ljung-Box

and Box-Pierce tests display power when they are applied on the residuals of the pooled point process and the marginal process of type- b events, although the power of the latter test dies down as the maximum lag M increases.

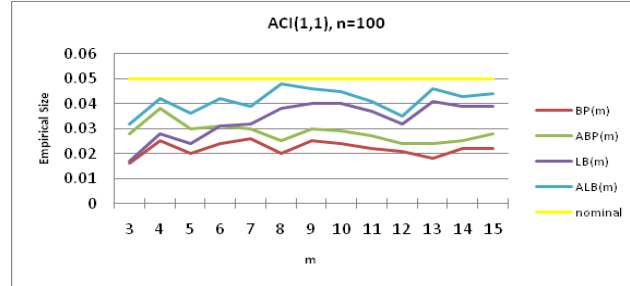


Figure 4.1: The empirical sizes of various portmanteau tests. DGP: $ACI(1,1)$, $df = M$, $n = 100$.

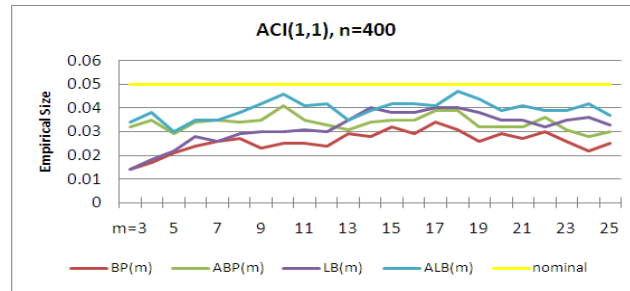


Figure 4.2: The empirical sizes of various portmanteau tests. DGP: $ACI(1,1)$, $df = M$, $n = 400$.

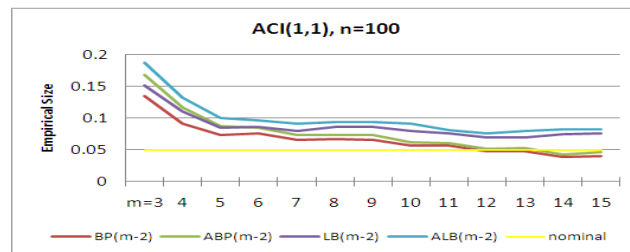


Figure 4.3: The empirical sizes of various portmanteau tests. DGP: $ACI(1,1)$, $df = M - 2$, $n = 100$.

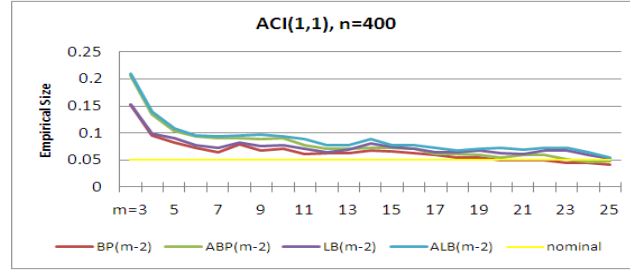


Figure 4.4: The empirical sizes of various portmanteau tests. DGP: $ACI(1,1)$, $df = M - 2$, $n = 400$.

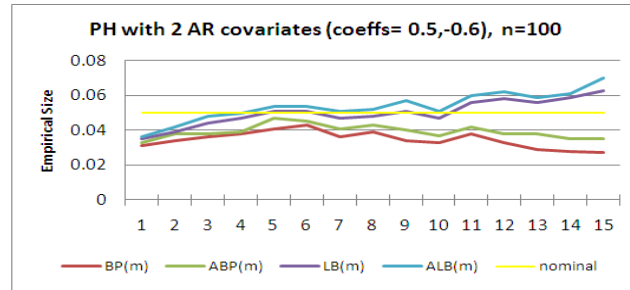


Figure 4.5: The empirical sizes of various portmanteau tests. DGP: Cox model, $df = M$, $n = 100$.

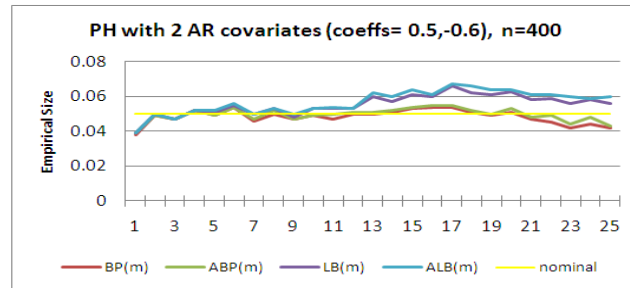


Figure 4.6: The empirical sizes of various portmanteau tests. DGP: Cox model, $df = M$, $n = 400$.

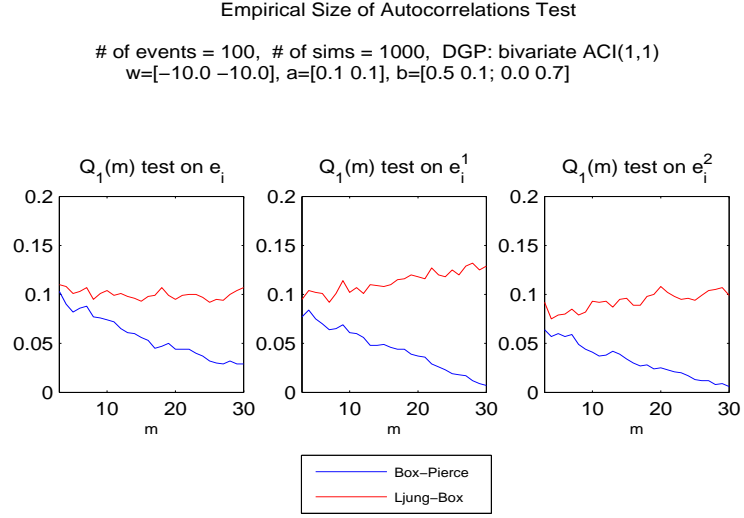


Figure 4.7: The empirical sizes of Box-Pierce and Ljung-Box tests. DGP: bivariate ACI(1,1) model, $df = M, n = 100$.

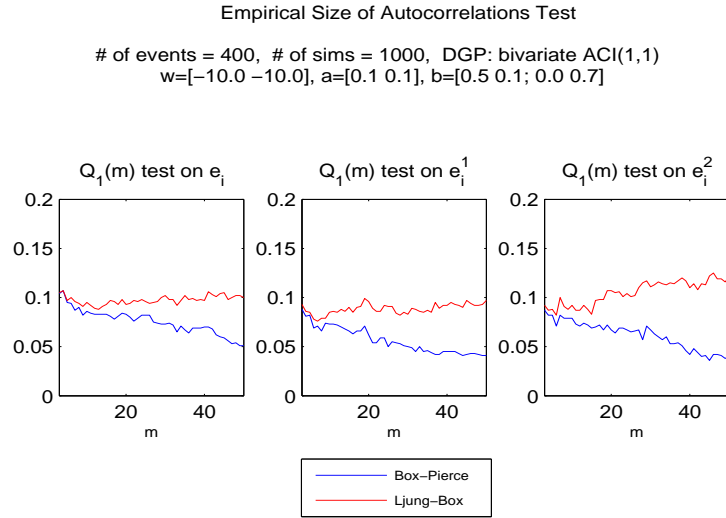


Figure 4.8: The empirical sizes of Box-Pierce and Ljung-Box tests. DGP: bivariate ACI(1,1) model, $df = M, n = 400$.

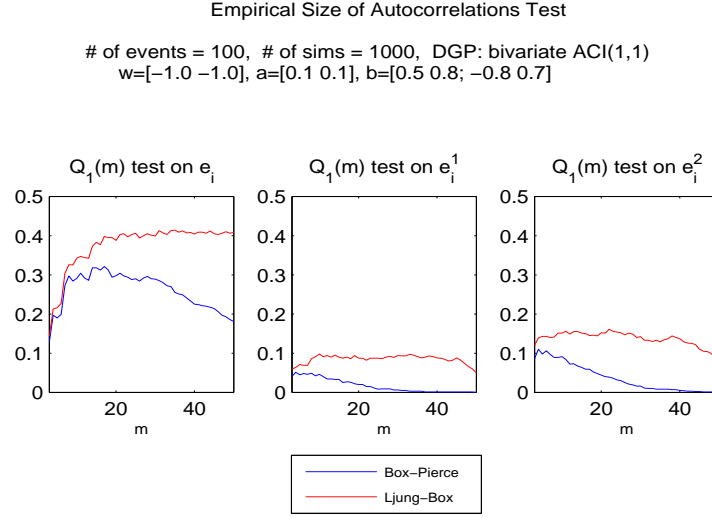


Figure 4.9: The empirical power of Box-Pierce and Ljung-Box tests. DGP: bivariate ACI(1,1) model, $df= M, n = 100$.

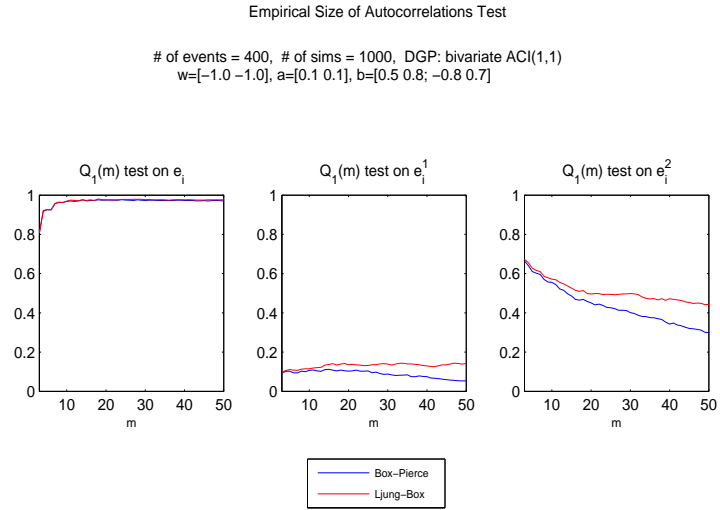


Figure 4.10: The empirical power of Box-Pierce and Ljung-Box tests. DGP: bivariate ACI(1,1) model, $df= M, n = 400$.

CHAPTER 5

CONCLUSION

Credit and financial contagion has been intriguing economists especially after the recent subprime mortgage crisis. Despite numerous studies surrounding this topic, the cause and mechanism of contagion are not yet clear to economists. This dissertation contributes to the literature by considering a number of econometric techniques to uncover the dynamics of credit and financial contagion. By transforming the problem into a meaningful Granger causality test, we can investigate the direction and magnitude of a contagion based on the observed point process data of credit events and negative shocks to financial markets. In particular, the nonparametric Granger causality test introduced in Chapter 2 offers a model-free approach for this purpose. For various bandwidth choice of the weighting function, this nonparametric test produces a profile of impulse response without the interference of a parametric model, and hence can be treated as a first-step analysis that facilitates the modelers to propose a parametric specification. After estimating the parametric model proposed by the modeler, a diagnostic checking procedure is necessary to evaluate the adequacy of the model. The portmanteau test considered in Chapter 4, the first of its kind for parametric multivariate conditional intensity models, is an indispensable technique in the model evaluation stage.

Applying these methodologies to the empirical study of Chapter 11 filings in the U.S., Chapter 3 provides empirical evidence that bankruptcy occurrences are contagious not only among firms in the same sector but also across different industries. Bankruptcy contagion tends to be the strongest and spread upstream along a typical supply chain near economic recessions and financial

crises. During the subprime mortgage crisis and its aftermath, there is a persistent spread of bankruptcy contagion from the financial sector to other sectors of the economy. This empirical evidence allows a more complete understanding of the nature of credit contagion and may facilitate policy makers, regulators and portfolio investors to make better decisions in preparation of the next financial crisis.

APPENDIX A

CHAPTER 2 APPENDIX

A.1 List of Assumptions

- (A1) The *pooled* counting process $N \equiv N^a + N^b$ is simple, i.e. $P(N(\{t\}) = 0 \text{ or } 1 \text{ for all } t) = 1$.
- (A2) The bivariate counting process $\mathbf{N} = (N^a, N^b)$ is *second-order stationary* and that the *second-order reduced product densities* $\varphi^{ij}(\cdot)$ ($i, j = a, b$) exist.
- (A3) The \mathcal{F} -conditional intensity $\lambda^k(t|\mathcal{F}_t^-)$ and \mathcal{F}^k -conditional intensity $\lambda_t^k \equiv \lambda^k(t|\mathcal{F}_t^k)$ of the counting process N_t^k exist and are predictable.
- (A4a) The kernel function $K(\cdot)$ is symmetric around zero and satisfies $\kappa_1 \equiv \int_{-\infty}^{\infty} K(u)du = 1$, $\kappa_2 \equiv \int_{-\infty}^{\infty} K^2(u)du < \infty$, $\kappa_4 \equiv \iiint_{(-\infty, \infty)} K(u)K(v)K(u+w)K(v+w)dudvdw < \infty$ and $\int_{-\infty}^{\infty} u^2 K(u)du < \infty$.
- (A4b) The kernel function $\mathring{K}(\cdot)$ is symmetric around zero and satisfies $\mathring{\kappa}_1 \equiv \int_{-\infty}^{\infty} \mathring{K}(u)du = 1$, $\mathring{\kappa}_2 \equiv \int_{-\infty}^{\infty} \mathring{K}^2(u)du < \infty$, $\mathring{\kappa}_4 \equiv \iiint_{(-\infty, \infty)} \mathring{K}(u)\mathring{K}(v)\mathring{K}(u+w)\mathring{K}(v+w)dudvdw < \infty$ and $\int_{-\infty}^{\infty} u^2 \mathring{K}(u)du < \infty$.
- (A4c) The kernel function $\ddot{K}(\cdot)$ is symmetric around zero and satisfies $\ddot{\kappa}_1 \equiv \int_{-\infty}^{\infty} \ddot{K}(u)du = 1$, $\ddot{\kappa}_2 \equiv \int_{-\infty}^{\infty} \ddot{K}^2(u)du < \infty$, $\ddot{\kappa}_4 \equiv \iiint_{(-\infty, \infty)} \ddot{K}(u)\ddot{K}(v)\ddot{K}(u+w)\ddot{K}(v+w)dudvdw < \infty$ and $\int_{-\infty}^{\infty} u^2 \ddot{K}(u)du < \infty$.
- (A4d) The kernels $K(x)$, $\mathring{K}(x)$ and $\ddot{K}(x)$ are all standard Gaussian kernels. That is: $K(x) = \mathring{K}(x) = \ddot{K}(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.
- (A5) The weighting function $w(\ell)$ is integrable over $(-\infty, \infty)$: i.e. $\int_{-\infty}^{\infty} w(\ell)d\ell < \infty$.
- (A6) $E\left[\{N^k(B_1)N^k(B_2)N^k(B_3)N^k(B_4)\}\right] < \infty$ for $k = a, b$ and for all bounded Borel sets B_i on \mathbb{R} , $i = 1, 2, 3, 4$.

- (A7) The rescaled counting process $\tilde{N}_u^k \equiv N_{Tu}^k/T$ (with natural filtration $\tilde{\mathcal{F}}^k$) has an $\tilde{\mathcal{F}}^k$ -conditional intensity function $\tilde{\lambda}_u^k$, which is twice continuously differentiable with respect to u , and is unobservable but deterministic.
- (A8) The joint cumulant $c_{22}(\ell_1, \ell_2, \ell_3)$ of $\{d\epsilon_s^a, d\epsilon_{s+\ell_1}^a, d\epsilon_{s+\ell_2}^b, d\epsilon_{s+\ell_3}^b\}$ is of order $o(a_T^2)$.

A.2 Figures

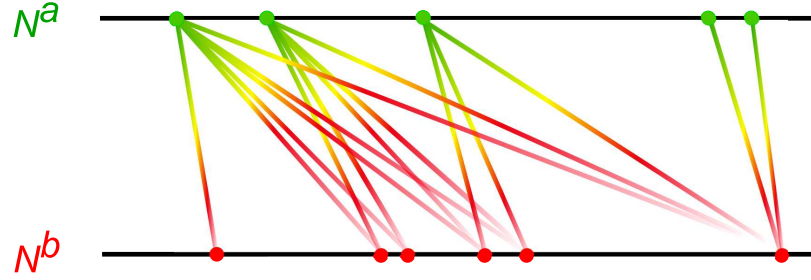
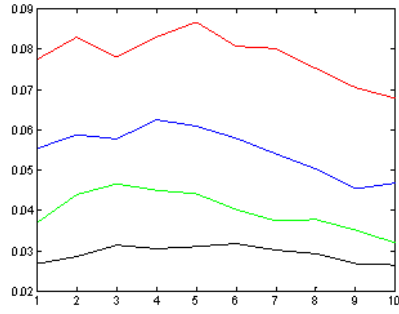
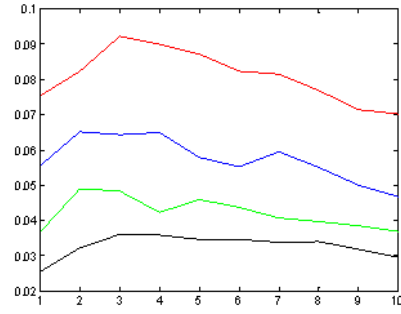


Figure A.1: The statistic Q aggregates the squared contributions of residual products $d\hat{\epsilon}_s^a d\hat{\epsilon}_t^b$ for all $s < t$. The lines join all pairs of type a and type b events (shocks) at their event times (τ_i^a, τ_j^a) for all $\tau_i^a < \tau_j^a$.

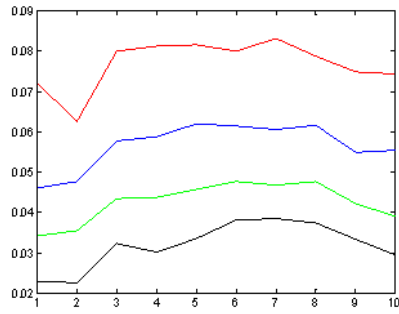
(a) $T = 500, B = 10, M = 60$



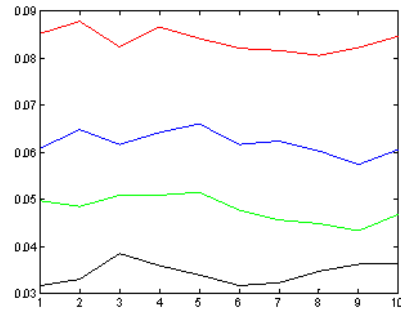
(b) $T = 1000, B = 10, M = 75$



(c) $T = 1500, B = 10, M = 100$



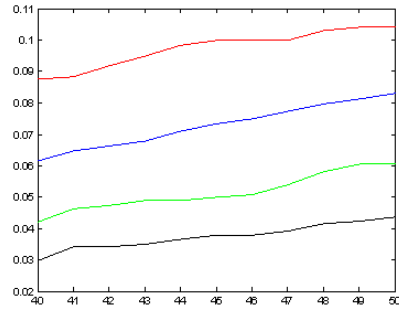
(d) $T = 2000, B = 10, M = 120$



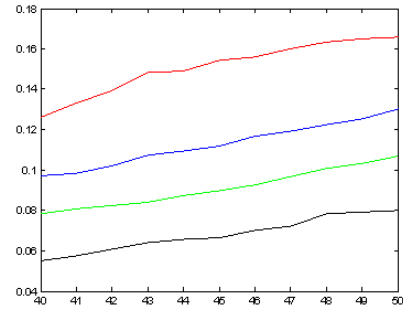
Plots of empirical rejection rates against B . Runs=5000, DGP= bivariate Poisson process (two independent Poisson processes with rate 0.1). Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.2: Size experiment of Q test, bivariate Poisson process.

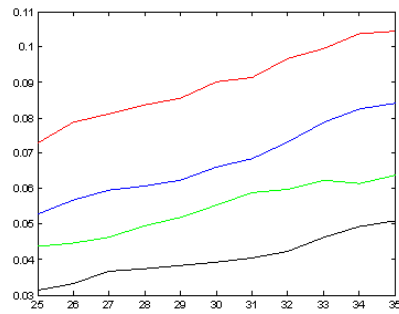
(a) size: $H = 3, B = 2$



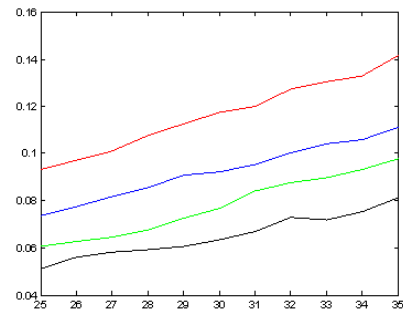
(b) power: $H = 3, B = 2$



(c) size: $H = 3, B = 20$



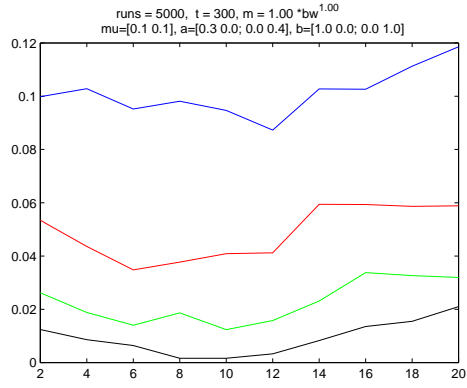
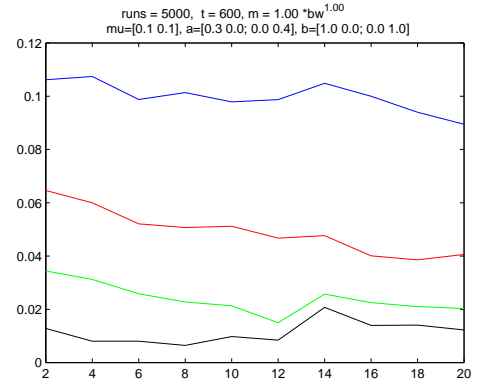
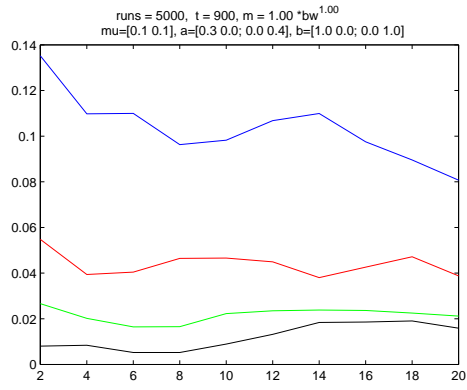
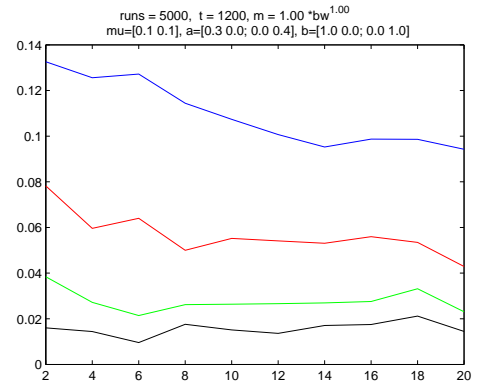
(d) power: $H = 3, B = 20$



Plots of em-

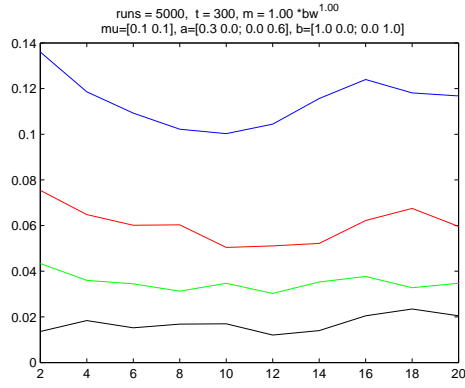
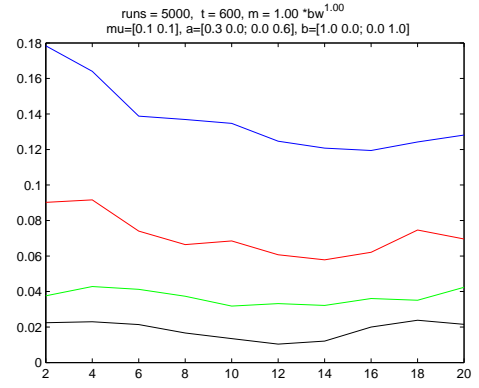
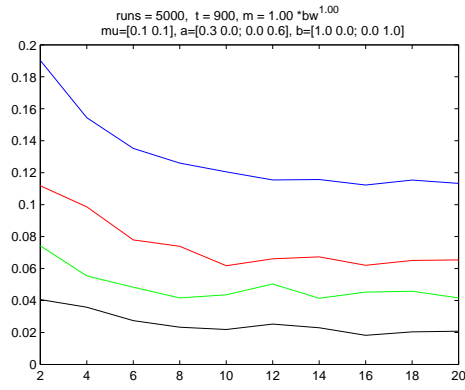
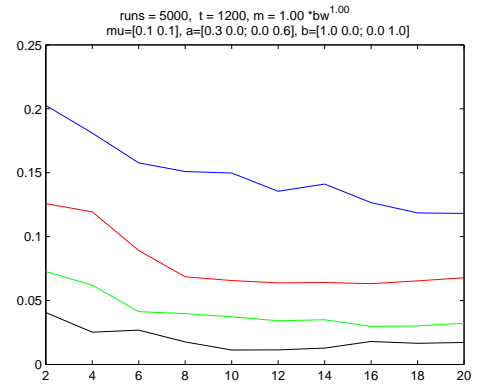
pirical rejection rates against B . Runs=10000, $T=1800$, DGP= bivariate exponential Hawkes model in (2.23). Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.3: Size and power experiment of Q test, bivariate exponential Hawkes process.

(a) $T = 300$ (b) $T = 600$ (c) $T = 900$ (d) $T = 1200$ 

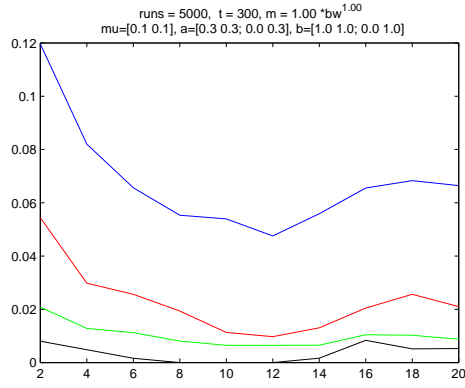
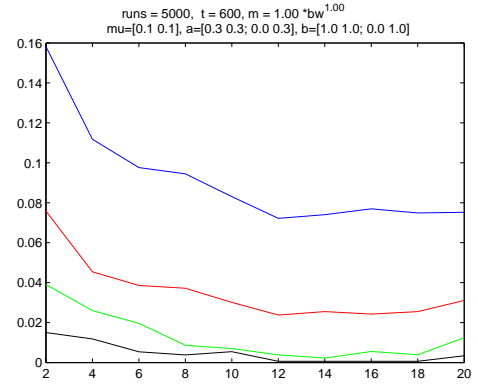
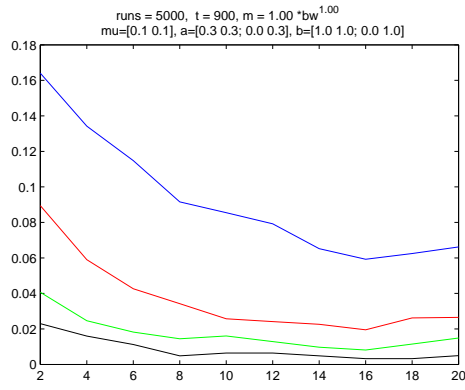
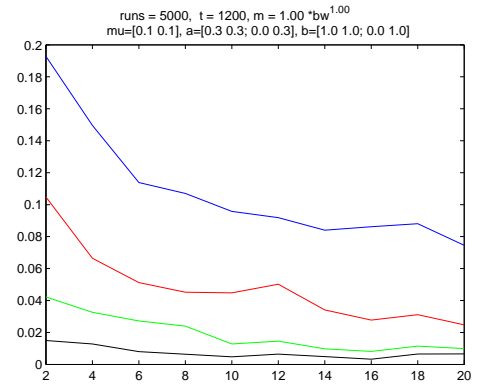
Plots of empirical rejection rates against B . Runs=5000, DGP=bivariate exponential Hawkes: $\mu = \begin{pmatrix} 0.1 & 0 \\ 0.1 & 0 \end{pmatrix}$, $\alpha = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.4 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.4: Size experiment 1 of Q^s test.

(a) $T = 300$ (b) $T = 600$ (c) $T = 900$ (d) $T = 1200$ 

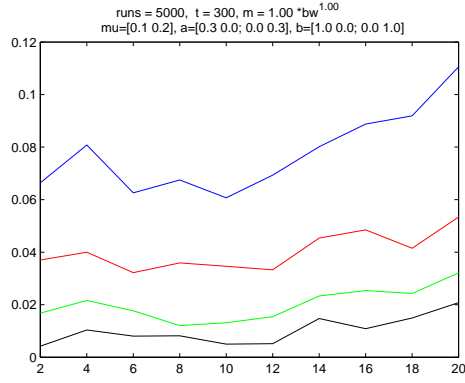
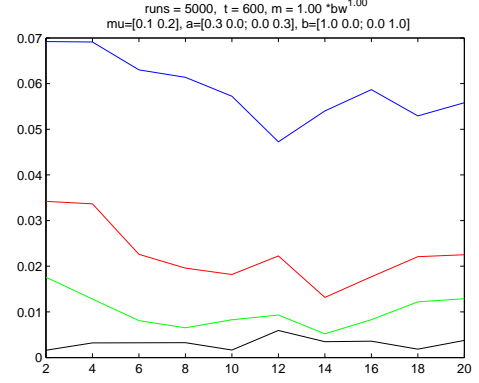
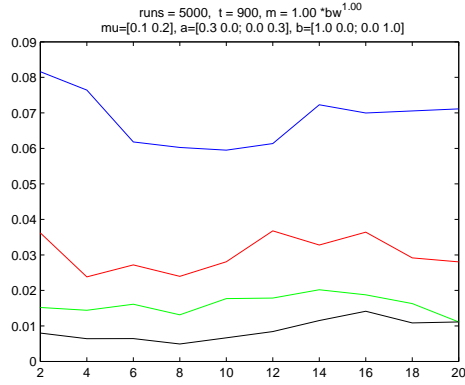
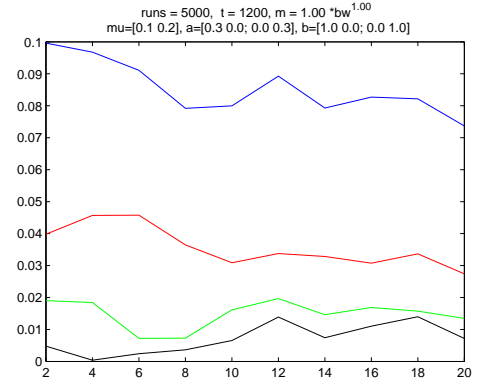
Plots of empirical rejection rates against B . Runs=5000, DGP= bivariate exponential Hawkes: $\mu = \begin{pmatrix} 0.1 & 0 \\ 0.1 & 0 \end{pmatrix}$, $\alpha = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.6 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.5: Size experiment 2 of Q^s test.

(a) $T = 300$ (b) $T = 600$ (c) $T = 900$ (d) $T = 1200$ 

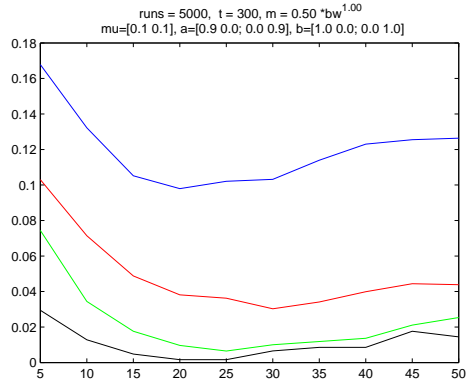
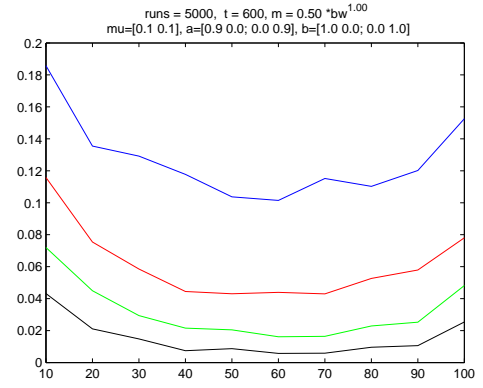
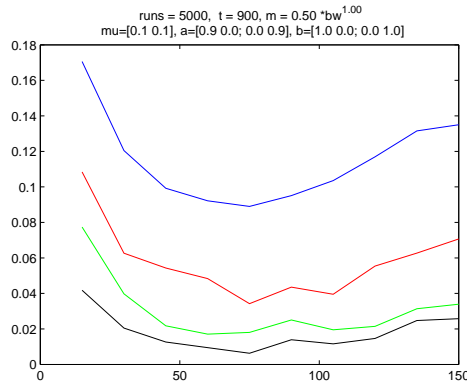
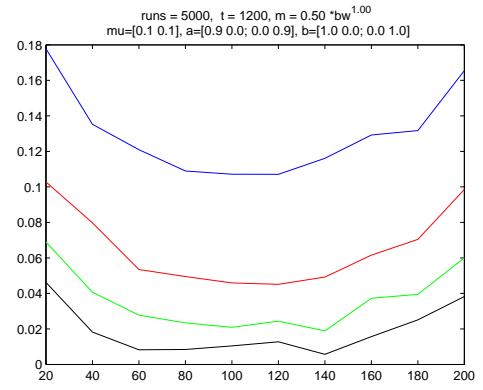
Plots of empirical rejection rates against B . Runs=5000, DGP= bivariate exponential Hawkes: $\mu = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}$, $\alpha = \begin{pmatrix} 0.3 & 0.3 \\ 0 & 0.3 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.6: Size experiment 3 of Q^s test.

(a) $T = 300$ (b) $T = 600$ (c) $T = 900$ (d) $T = 1200$ 

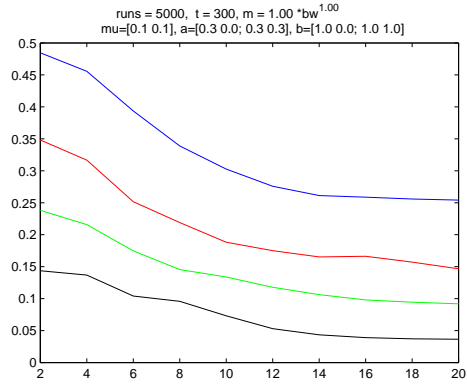
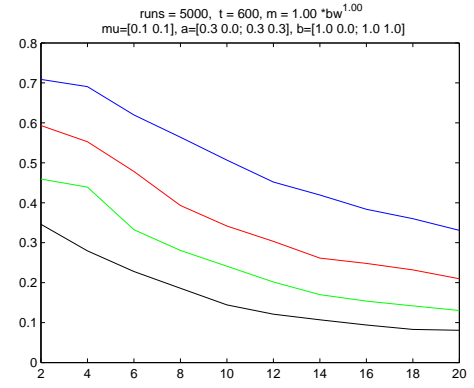
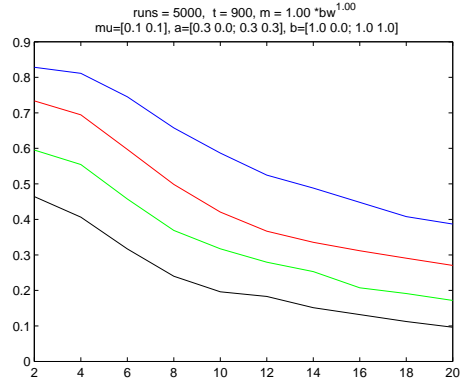
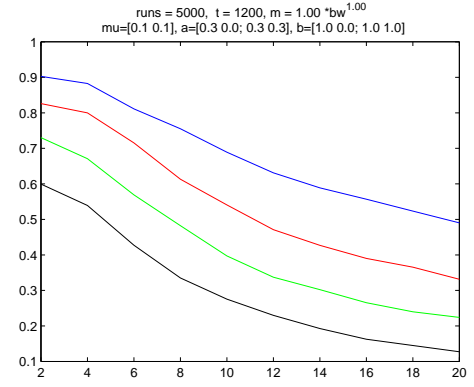
Plots of empirical rejection rates against B . Runs=5000, DGP= bivariate exponential Hawkes: $\mu = \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.1 \end{pmatrix}$, $\alpha = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.3 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.7: Size experiment 4 of Q^s test.

(a) $T = 300$ (b) $T = 600$ (c) $T = 900$ (d) $T = 1200$ 

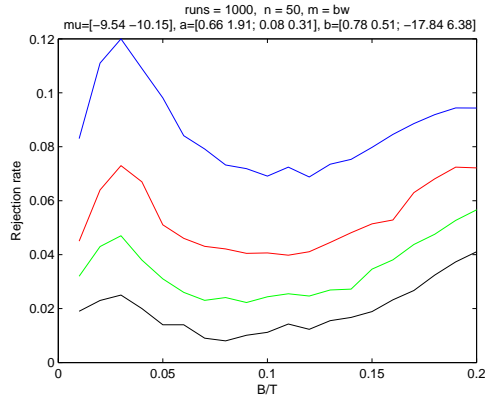
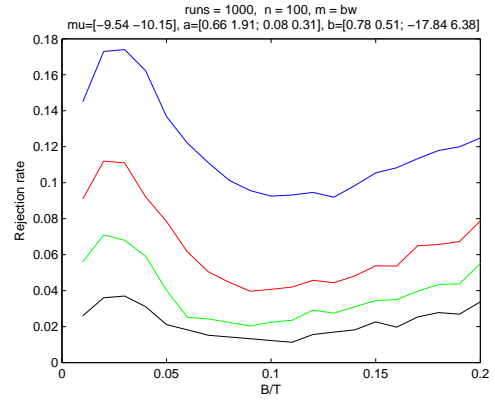
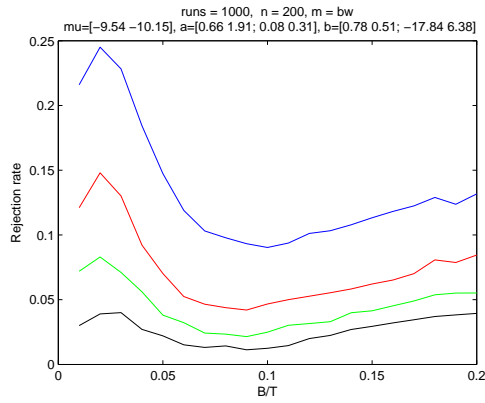
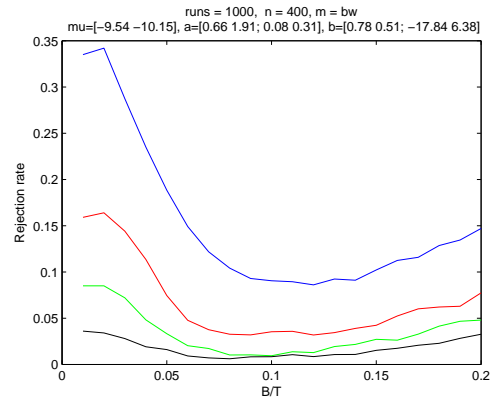
Plots of empirical rejection rates against B . Runs=5000, DGP= bivariate exponential Hawkes: $\mu = \begin{pmatrix} 0.1 & 0 \\ 0.1 & 0 \end{pmatrix}$, $\alpha = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.8: Size experiment 5 of Q^s test.

(a) $T = 300$ (b) $T = 600$ (c) $T = 900$ (d) $T = 1200$ 

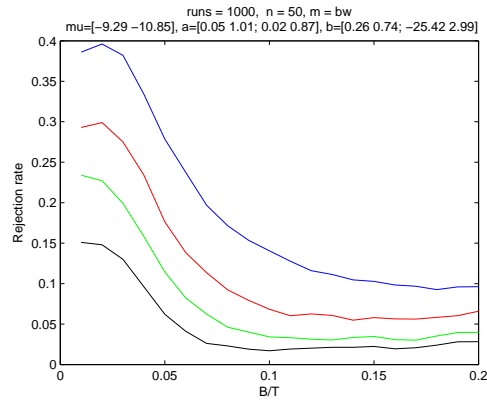
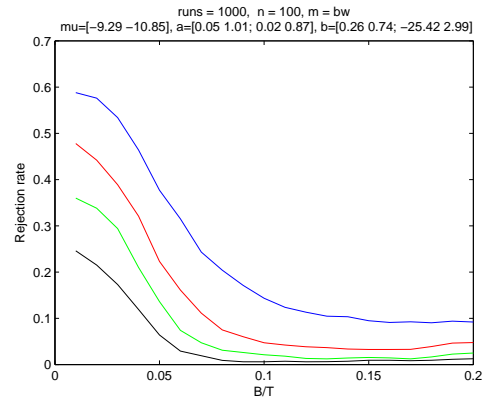
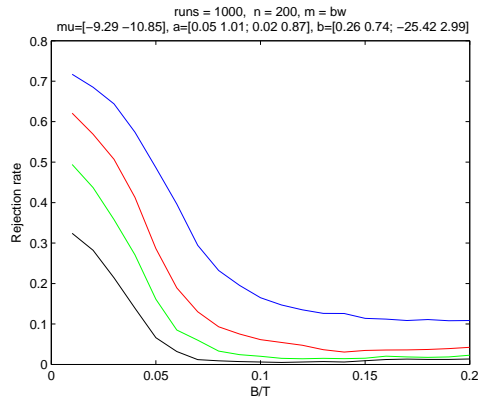
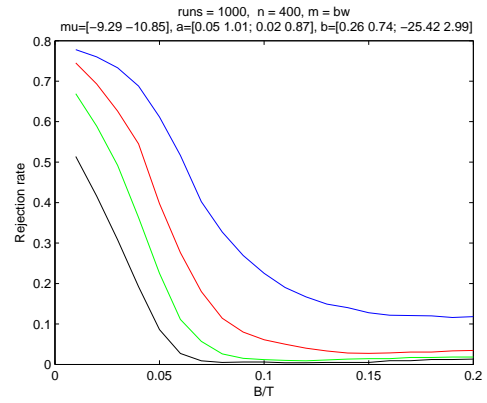
Plots of empirical rejection rates against B . Runs=5000, DGP= bivariate exponential Hawkes: $\mu = \begin{pmatrix} 0.1 & 0 \\ 0.1 & 0 \end{pmatrix}$, $\alpha = \begin{pmatrix} 0.3 & 0 \\ 0.3 & 0.3 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.9: Power experiment of Q^s test.

(a) $N(T) = 50$ (b) $N(T) = 100$ (c) $N(T) = 200$ (d) $N(T) = 400$ 

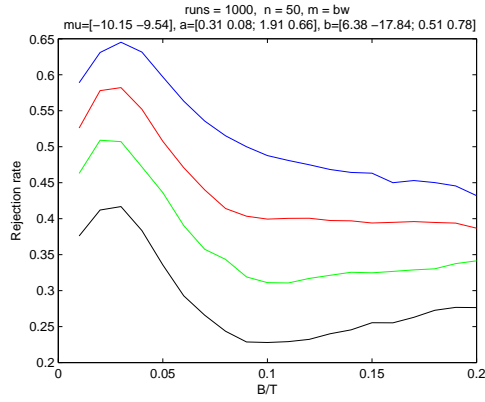
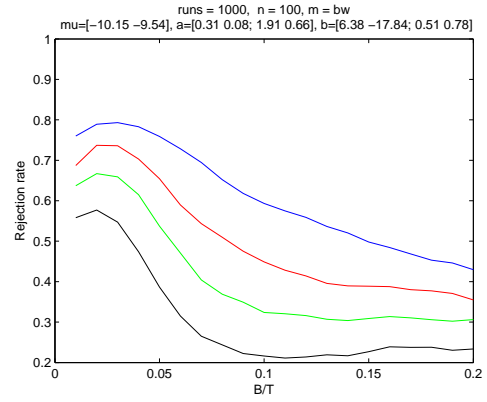
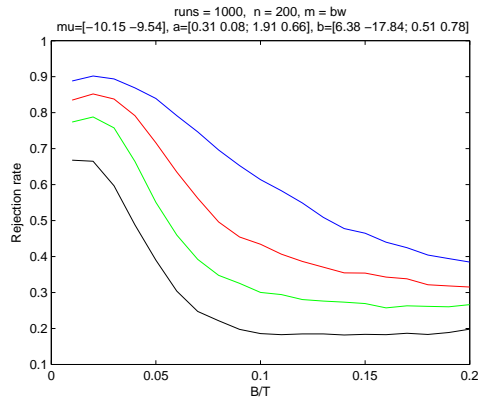
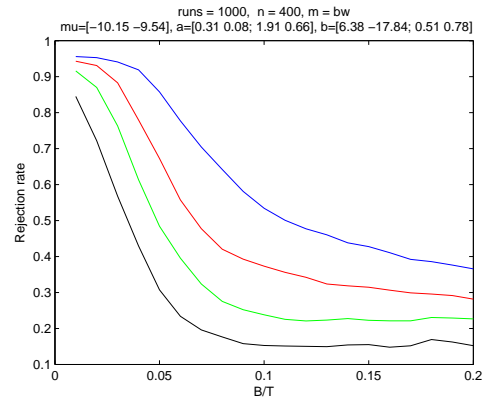
Plots of empirical rejection rates against B/T . Runs=1000, $N(T)$ = total number of events in the pooled point process, DGP = bivariate ACI(1,1) model estimated from bankruptcy data (upstream and downstream, Mar 2001 - Nov 2001). Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.10: Size experiment R-S1 of Q^s test.

(a) $N(T) = 50$ (b) $N(T) = 100$ (c) $N(T) = 200$ (d) $N(T) = 400$ 

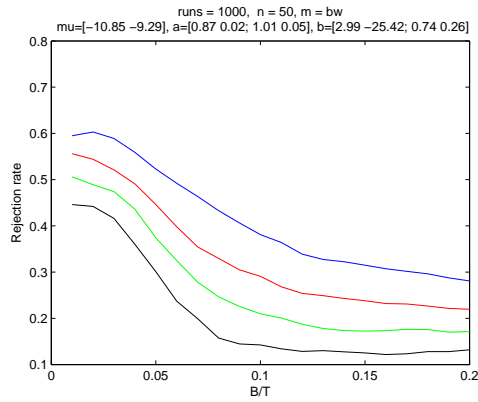
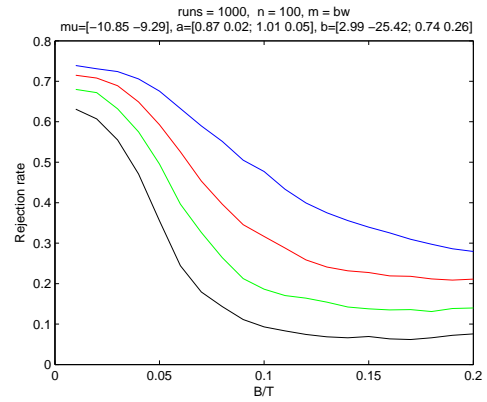
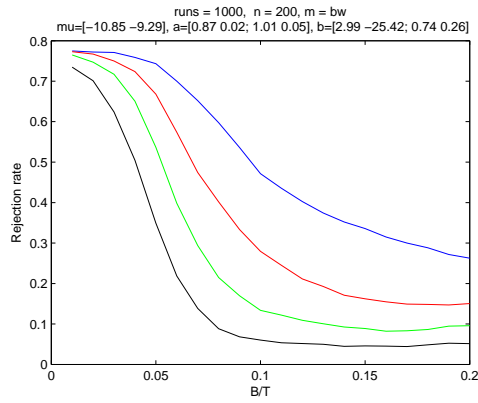
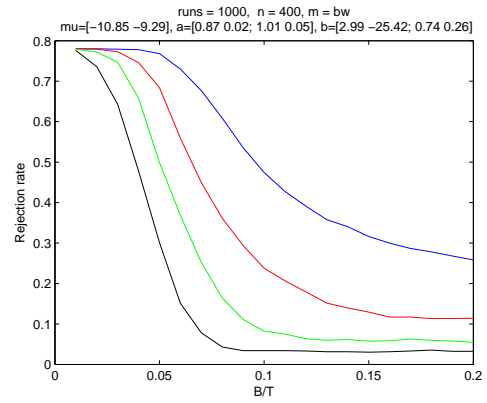
Plots of empirical rejection rates against B/T . Runs=1000, $N(T)$ = total number of events in the pooled point process, DGP = bivariate ACI(1,1) model estimated from bankruptcy data (upstream and downstream, Jul 2009 - Nov 2011). Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.11: Size experiment R-S2 of Q^s test.

(a) $N(T) = 50$ (b) $N(T) = 100$ (c) $N(T) = 200$ (d) $N(T) = 400$ 

Plots of empirical rejection rates against B/T . Runs=1000, $N(T)$ = total number of events in the pooled point process, DGP = bivariate ACI(1,1) model estimated from bankruptcy data (upstream and downstream, Mar 2001 - Nov 2001). Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

Figure A.12: Power experiment R-P1 of Q^s test.

(a) $N(T) = 50$ (b) $N(T) = 100$ (c) $N(T) = 200$ (d) $N(T) = 400$ 

Plots of empirical rejection rates against B/T . Runs=1000, $N(T)$ = total number of events in the pooled point process, DGP = bivariate ACI(1,1) model estimated from bankruptcy data (upstream and downstream, Jul 2009 - Nov 2011). Nominal size: blue=0.1; red=0.05; green=0.025; black=0.01.

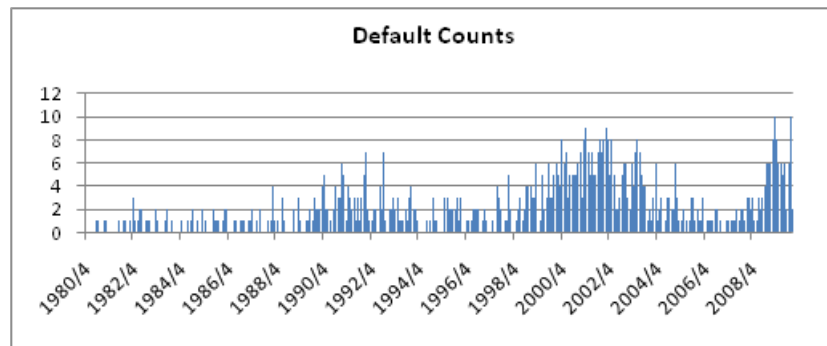
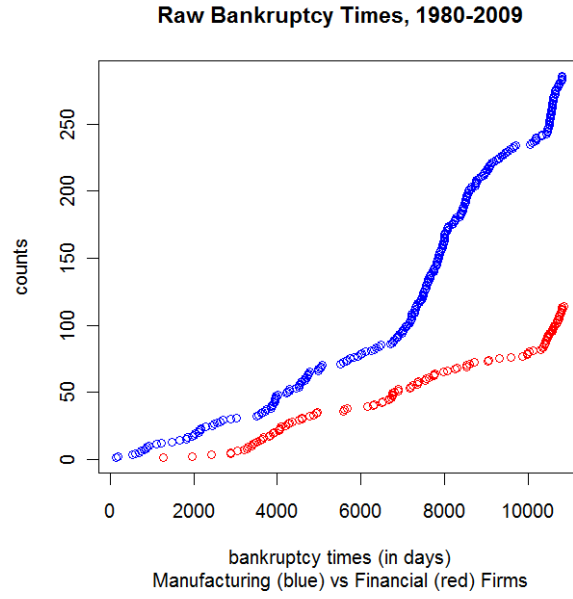
Figure A.13: Power experiment R-P2 of Q^s test.

Figure A.14: Histogram of bankruptcies of U.S. firms, 1980–2010.



N_t^m (Blue): A: Agricultural; B: Mining; C: Construction; D: Manufacturing; E: Transportation, Communications, Electric, Gas; N_t^f (Red): F: Wholesale; G: Retail; H: Finance, Insurance, Real Estate; I: Services

Figure A.15: Raw counts of bankruptcies in manufacturing and financial related sectors.

A.3 Tables

Scheme	Observation window	Sample size	Limit	Duration
1	$[0, T]$	$n = N(T)$	$T \rightarrow \infty \Rightarrow n \rightarrow \infty$	$\tau_i - \tau_{i-1}$ fixed
2	$[0, T_0]$	$n = N(T_0)$	$n \rightarrow \infty, T_0$ fixed	$\tau_i - \tau_{i-1} \downarrow 0$

Table A.1: The asymptotic mechanisms of the two schemes.

H	B	M	Trade \rightarrow Quote			Quote \rightarrow Trade		
sig. levels:			0.1	0.05	0.01	0.1	0.05	0.01
0.6	2	20	3	1	1	4	2	1
0.6	4	17	4	3	2	4	2	1
0.6	10	15	7	5	3	1	1	0
0.6	20	10	4	2	1	1	1	1
1	2	38	8	7	3	6	6	3
1	4	35	9	6	3	4	3	3
1	10	30	15	15	15	4	3	2
1	20	27	16	15	11	3	2	2
3	2	40	8	6	5	7	6	3
3	4	35	13	11	6	8	7	2
3	10	33	19	16	11	7	5	4
3	20	30	15	12	11	4	2	2

Mean number of trades=88.8, quotes=325.1. The bandwidth combinations give right sizes in simulations (with an estimated bivariate Hawkes model to PG data as DGP). Bandwidths (in days) of (i) cross-covariance function: H ; (ii) weighting function: B ; (iii) conditional intensity: M . All kernels are Gaussian.

Table A.2: Significant day counts (out of 41 days) of PG, 9:45am – 10:15am.

H	B	M	Trade \rightarrow Quote			Quote \rightarrow Trade		
sig. levels:			0.1	0.05	0.01	0.1	0.05	0.01
0.6	2	20	11	8	6	8	6	4
0.6	4	17	16	15	11	9	7	4
0.6	10	15	17	16	15	8	6	5
0.6	20	10	9	6	6	7	5	3
1	2	38	6	5	4	8	8	6
1	4	35	20	18	13	10	8	7
1	10	30	17	16	13	11	10	8
1	20	27	17	16	14	13	11	9
3	2	40	14	9	4	17	12	6
3	4	35	24	20	18	19	14	11
3	10	33	26	25	22	24	20	18
3	20	30	25	23	18	26	25	16

Mean number of trades=103.8, quotes=403.73. The bandwidth combinations give right sizes in simulations (with an estimated bivariate Hawkes model to PG data as DGP). Bandwidths (in days) of (i) cross-covariance function: H ; (ii) weighting function: B ; (iii) conditional intensity: M . All kernels are Gaussian.

Table A.3: Significant day counts (out of 41 days) of PG, 11:45am – 12:45pm.

H	B	M	Trade \rightarrow Quote			Quote \rightarrow Trade		
sig. levels:			0.1	0.05	0.01	0.1	0.05	0.01
0.6	2	20	1	0	0	2	1	1
0.6	4	17	7	5	3	1	1	0
0.6	10	15	8	7	7	0	0	0
0.6	20	10	6	5	3	1	1	0
1	2	38	4	3	2	4	3	1
1	4	35	5	5	4	2	2	1
1	10	30	18	18	16	6	5	2
1	20	27	13	13	13	2	2	0
3	2	40	5	5	3	6	4	3
3	4	35	10	9	7	6	6	4
3	10	33	14	13	12	7	7	5
3	20	30	10	10	9	8	6	2

Mean number of trades=93.7, quotes=361.56. The bandwidth combinations give right sizes in simulations (with an estimated bivariate Hawkes model to PG data as DGP). Bandwidths (in days) of (i) cross-covariance function: H ; (ii) weighting function: B ; (iii) conditional intensity: M . All kernels are Gaussian.

Table A.4: Significant day counts (out of 41 days) of PG, 3:30pm – 4:00pm.

PG periods	B sig.:	Trade \rightarrow Quote			Quote \rightarrow Trade		
		0.1	0.05	0.01	0.1	0.05	0.01
09:45-10:15	10	13	8	0	1	0	0
	20	24	13	4	3	1	0
$\mu^t = 88.8$	30	23	14	4	2	1	0
$\mu^q = 325.1$	40	21	16	4	1	1	0
11:45-12:45	10	32	28	16	2	0	0
	20	35	34	21	3	0	0
$\mu^t = 103.8$	30	34	34	16	5	0	0
$\mu^q = 403.7$	40	33	30	14	6	0	0
15:30-16:00	10	26	12	3	0	0	0
	20	30	21	3	1	0	0
$\mu^t = 93.7$	30	26	16	6	3	1	0
$\mu^q = 361.6$	40	20	11	4	3	0	0

μ^t =mean number of trades, μ^q =mean number of quotes. The bandwidths R^k of unconditional autocorrelation estimators $\hat{c}_{R^k}^{kk}(\theta)$ (for $k = trade$ and $quote$) are set equal to B , the bandwidth of the weighting function $w_B(\theta)$.

Table A.5: Significant day counts (out of 41 days) of PG over various trading hours of a day.

GM periods	B sig.:	Trade \rightarrow Quote			Quote \rightarrow Trade		
		0.1	0.05	0.01	0.1	0.05	0.01
09:45-10:15	10	6	1	0	0	0	0
	20	13	8	0	1	1	0
$\mu^t = 65.4$	30	15	11	1	2	2	0
$\mu^q = 191.9$	40	17	11	4	2	2	0
11:45-12:45	10	26	16	6	9	2	1
	20	28	19	10	10	4	0
$\mu^t = 80.5$	30	26	18	8	7	2	0
$\mu^q = 217.3$	40	24	20	7	9	2	0
15:30-16:00	10	8	4	0	2	1	0
	20	12	7	1	4	1	0
$\mu^t = 65.1$	30	11	5	0	5	3	0
$\mu^q = 188.9$	40	10	5	0	6	2	1

μ^t =mean number of trades, μ^q =mean number of quotes. The bandwidths R^k of unconditional autocorrelation estimators $\hat{c}_{R^k}^{kk}(\theta)$ (for $k = trade$ and $quote$) are set equal to B , the bandwidth of the weighting function $w_B(\theta)$.

Table A.6: Significant day counts (out of 41 days) of GM over various trading hours of a day.

H	$B = M = 365$		$B = M = 548$		$B = M = 730$	
	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$
2	2.32	3.56	0.09	12.66	-0.90	31.81
4	-3.85	5.87	-4.01	17.58	0.90	43.21
6	-3.14	-0.45	-2.21	10.87	5.97	49.86
8	-2.21	0.86	-0.82	15.22	9.21	49.06
10	-1.53	1.93	0.15	18.63	11.52	57.37
12	-1.06	2.62	0.86	21.12	13.32	63.88
14	-0.67	3.04	1.41	22.93	14.73	69.04

Sample sizes: $(n^m, n^f) = (209, 149)$. $m \rightarrow f$ ($f \rightarrow m$) denotes bankruptcy contagion from manufacturing related to financial related firms (and vice versa). One-sided critical values: $z_{0.05} = 1.64$; $z_{0.01} = 2.33$. Bandwidths (in days) of (i) cross-covariance function: H ; (ii) weighting function: B ; (iii) conditional intensity: M . All kernels are Gaussian.

Table A.7: Q tests on bankruptcy data, Sep96 – Jul03.

H	$B = M = 365$		$B = M = 548$		$B = M = 730$	
	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$
2	-5.12	0.55	-4.56	-1.21	-2.98	-1.77
4	-3.13	1.77	-2.24	-0.24	0.01	-0.71
6	-2.70	1.17	-1.38	-0.34	1.46	-0.21
8	-1.96	0.32	-0.21	-0.68	3.17	0.04
10	-1.14	-0.07	0.96	-0.78	4.77	0.35
12	-0.46	-0.09	1.88	-0.65	5.95	0.75
14	0.04	0.10	2.53	-0.41	6.72	1.17

Sample sizes: $(n^m, n^f) = (65, 29)$. $m \rightarrow f$ ($f \rightarrow m$) denotes bankruptcy contagion from manufacturing related to financial related firms (and vice versa). One-sided critical values: $z_{0.05} = 1.64$; $z_{0.01} = 2.33$. Bandwidths (in days) of (i) cross-covariance function: H ; (ii) weighting function: B ; (iii) conditional intensity: M . All kernels are Gaussian.

Table A.8: Q test on bankruptcy data, Aug03 – Aug07.

H	$B = M = 365$		$B = M = 548$		$B = M = 730$	
	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$	$J^{m \rightarrow f}$	$J^{f \rightarrow m}$
2	19.37	7.58	50.42	28.11	83.53	52.44
4	10.25	14.67	46.67	39.19	89.14	73.09
6	12.80	17.67	56.57	56.38	108.40	100.61
8	15.37	20.86	65.69	65.72	125.76	117.12
10	22.14	23.29	73.90	73.13	141.37	130.41
12	23.24	25.23	81.46	79.42	155.63	141.87
14	24.43	26.87	93.37	85.00	175.19	152.16

Sample sizes: $(n^m, n^f) = (78, 71)$. $m \rightarrow f$ ($f \rightarrow m$) denotes bankruptcy contagion from manufacturing related to financial related firms (and vice versa). One-sided critical values: $z_{0.05} = 1.64$; $z_{0.01} = 2.33$. Bandwidths (in days) of (i) cross-covariance function: H ; (ii) weighting function: B ; (iii) conditional intensity: M . All kernels are Gaussian.

Table A.9: Q tests on bankruptcy data, Sep07 – Jun10.

B	$m \rightarrow f$	p-value	$f \rightarrow m$	p-value
30	1.79	0.037	1.78	0.038
60	1.74	0.041	1.93	0.027
90	1.66	0.049	1.99	0.024
120	1.58	0.057	1.99	0.024
150	1.51	0.066	1.96	0.025
180	1.43	0.076	1.93	0.027
210	1.35	0.088	1.89	0.029
240	1.27	0.103	1.86	0.032
270	1.18	0.119	1.82	0.034
300	1.09	0.138	1.79	0.037

$m \rightarrow f$ ($f \rightarrow m$) denotes bankruptcy contagion from manufacturing related to financial related firms (and vice versa). One-sided critical values: $z_{0.05} = 1.64$; $z_{0.01} = 2.33$. Bandwidth (in days) of the weighting function: B . Bandwidths R^k of autocovariance function estimators are set equal to 300. All kernels are Gaussian.

Table A.10: Q^s test on bankruptcy data, September 1996 – June 2010.

Index	Trading hours (local time)	GMT	Start date
DJI	09:30 - 16:00	-5	10/1/1928
FTSE	08:00 - 16:30	+0	4/2/1984
DAX	09:00 - 17:30	+1	11/26/1990
CAC	09:00 - 17:30	+1	3/1/1990
HSI	10:00 - 12:30, 14:30 - 16:00 ¹	+8	12/31/1986
STI	09:00 - 12:30, 2:00 - 5:00	+8	12/28/1987
TWI	09:00 - 13:30	+8	7/2/1997
NIK	09:00 - 11:00, 12:30 - 15:00	+9	1/4/1984
AOI	10:00 - 16:00	+10	8/3/1984

Adjusted daily index values were collected from Yahoo! Finance. The end date of all the time series is 8/19/2011. Each time, a Granger causality test is performed on the event sequences of a pair of indices, with the sampling period equal to the shorter of the two sampling periods of the two indices.

Table A.11: Trading hours, Greenwich mean time and start dates of the sampling periods of major stock indices.

B	90% VaR		95% VaR		99% VaR	
	HSI→DJI	DJI→HSI	HSI→DJI	DJI→HSI	HSI→DJI	DJI→HSI
1	0.92	1.69	1.28	3.00	7.35	9.61
2	0.75	1.17	1.04	2.16	5.98	7.92
3	0.67	0.95	0.98	1.81	5.23	6.99
5	0.60	0.76	0.96	1.49	4.63	6.44
10	0.53	0.60	0.96	1.19	4.06	5.85
(n_1, n_2)	(608,620)		(304,310)		(61,62)	

The bandwidths of the autocovariance functions are chosen to be $R^k = 10.5B^{0.3}$.

Table A.12: Q^s test applied to extreme negative shocks of DJI and HSI.

B	90% VaR		95% VaR		99% VaR	
	NIK→DJI	DJI→NIK	NIK→DJI	DJI→NIK	NIK→DJI	DJI→NIK
1	0.56	1.64	1.23	2.87	5.70	10.62
2	0.46	1.13	1.00	2.05	5.21	7.84
3	0.43	0.93	0.90	1.72	5.12	7.01
5	0.41	0.74	0.82	1.42	4.93	6.52
10	0.38	0.57	0.74	1.11	4.49	6.18
(n_1, n_2)	(604,614)		(303,306)		(63,59)	

The bandwidths of the autocovariance functions are chosen to be $R^k = 10.5B^{0.3}$.

Table A.13: Q^s test applied to extreme negative shocks of DJI and NIK.

B	90% VaR		95% VaR		99% VaR	
	FTS→DJI	DJI→FTS	FTS→DJI	DJI→FTS	FTS→DJI	DJI→FTS
1	2.88	0.88	4.93	1.76	18.53	5.81
2	1.82	0.86	3.18	1.74	13.25	6.35
3	1.46	0.81	2.59	1.68	11.25	6.38
5	1.16	0.76	2.07	1.62	9.16	6.46
10	0.90	0.68	1.65	1.45	7.31	6.56
(n_1, n_2)	(621,620)		(311,310)		(63,62)	

The bandwidths of the autocovariance functions are chosen to be $R^k = 10.5B^{0.3}$.

Table A.14: Q^s test applied to extreme negative shocks of DJI and FTSE.

B	90% VaR		95% VaR		99% VaR	
	AOI→DJI	DJI→AOI	AOI→DJI	DJI→AOI	AOI→DJI	DJI→AOI
1	0.59	2.32	1.37	4.72	7.10	14.28
2	0.60	1.48	1.29	3.19	6.80	10.82
3	0.60	1.16	1.25	2.56	6.39	9.58
5	0.57	0.89	1.17	1.98	5.57	8.59
10	0.50	0.67	1.08	1.47	4.74	7.82
(n_1, n_2)	(679,680)		(340,341)		(68,69)	

The bandwidths of the autocovariance functions are chosen to be $R^k = 10.5B^{0.3}$.

Table A.15: Q^s test applied to extreme negative shocks of DJI and AOI.

A.4 Proof of Theorem 5

I first expand (2.12) into

$$\begin{aligned}
\hat{\gamma}_H(\ell) &= \frac{1}{TH} \int_0^T \int_0^T K\left(\frac{t-s-\ell}{H}\right) (dN_s^a - \hat{\lambda}_s^a ds) (dN_t^b - \hat{\lambda}_t^b dt) \\
&= \frac{1}{TH} \int_0^T \int_0^T K\left(\frac{t-s-\ell}{H}\right) [dN_s^a dN_t^b - \hat{\lambda}_s^a ds dN_t^b - dN_s^a \hat{\lambda}_t^b dt + \hat{\lambda}_s^a \hat{\lambda}_t^b ds dt] \\
&= A_1 + A_2 + A_3 + A_4.
\end{aligned} \tag{A.1}$$

The first term is

$$A_1 = \frac{1}{TH} \sum_{i=1}^{N_T^a} \sum_{j=1}^{N_T^b} K\left(\frac{t_j^b - t_i^a - \ell}{H}\right).$$

The second term, after substituting $\hat{\lambda}_t^b$ by (2.14), becomes

$$A_2 = -\frac{1}{THM} \sum_{i=1}^{N_T^a} \sum_{j=1}^{N_T^b} \int_0^T K\left(\frac{t_j^b - s - \ell}{H}\right) \mathring{K}\left(\frac{s - t_i^a}{M}\right) ds.$$

Similarly, the third term is

$$A_3 = -\frac{1}{THM} \sum_{i=1}^{N_T^a} \sum_{j=1}^{N_T^b} \int_0^T K\left(\frac{t - t_i^a - \ell}{H}\right) \mathring{K}\left(\frac{t_j^b - t}{M}\right) dt,$$

and the fourth one is

$$A_4 = \frac{1}{THM^2} \sum_{i=1}^{N_T^a} \sum_{j=1}^{N_T^b} \int_0^T \int_0^T K\left(\frac{t-s-\ell}{H}\right) \mathring{K}\left(\frac{s-t_i^a}{M}\right) \mathring{K}\left(\frac{t_j^b-t}{M}\right) ds dt.$$

Note that the last three terms involve the convolution of the kernels $K(\cdot)$ and $\mathring{K}(\cdot)$ (twice for A_4).

Under assumption (A4d), I can simplify the expressions further, as it is well known that Gaussian kernels are invariant under convolution: for any $H_1, H_2 > 0$, the Gaussian kernel $K(\cdot)$ enjoys the property that

$$\frac{1}{H_1 H_2} \int_{-\infty}^{\infty} K\left(\frac{x-z}{H_1}\right) K\left(\frac{z}{H_2}\right) dz = \frac{1}{\sqrt{H_1^2 + H_2^2}} K\left(\frac{x}{\sqrt{H_1^2 + H_2^2}}\right).$$

Using this invariance property and a change of variables, I can simplify the integrations and rewrite (A.1) as

$$\hat{\gamma}_H(\ell) = \frac{1}{T} \sum_{i=1}^{N_T^a} \sum_{j=1}^{N_T^b} \left[\frac{1}{H} K\left(\frac{t_j^b - t_i^a - \ell}{H}\right) - \frac{2}{\sqrt{H^2 + M^2}} K\left(\frac{t_j^b - t_i^a - \ell}{\sqrt{H^2 + M^2}}\right) + \frac{1}{\sqrt{H^2 + 2M^2}} K\left(\frac{t_j^b - t_i^a - \ell}{\sqrt{H^2 + 2M^2}}\right) \right].$$

A.5 Proof of Theorem 6

Let us prove asymptotic normality first. For notational convenience, I drop the superscript k of N_t^k , λ_t^k and their rescaled version in this proof. Let $\lambda_t^* = \int_0^T \frac{1}{M} \dot{K}\left(\frac{t-s}{M}\right) \lambda_s ds$, then I can rewrite (2.18) into

$$\sqrt{M} \left(\frac{\hat{\lambda}_{T_v} - \lambda_{T_v}^*}{\sqrt{\lambda_{T_v}}} \right) + \sqrt{M} \left(\frac{\lambda_{T_v}^* - \lambda_{T_v}}{\sqrt{\lambda_{T_v}}} \right) =: X_1 + X_2. \quad (\text{A.2})$$

Suppose $\tilde{\mathcal{F}}^k$ denotes the natural filtration of the rescaled counting process \tilde{N}^k . Then, it follows from (2.16) that $\tilde{\lambda}_u = \lambda_{T_u}$.

With a change of variables $t = Tv$ and $s = Tu$, the first term of (A.2) becomes

$$\begin{aligned} X_1 &= \int_0^1 \frac{1}{\sqrt{M}} \dot{K}\left(\frac{v-u}{M/T}\right) \left(\frac{dN_{T_u} - \lambda_{T_u} T du}{\sqrt{\lambda_{T_v}}} \right) \\ &= \int_0^1 \frac{1}{\sqrt{M}} \dot{K}\left(\frac{v-u}{M/T}\right) \left(\frac{d\tilde{N}_u - \tilde{\lambda}_u du}{\sqrt{\tilde{\lambda}_v/T}} \right). \end{aligned}$$

The multiplicative model in Ramlau-Hansen (R-H, 1983) assumes that $\tilde{\lambda}_u \equiv \tilde{\lambda}_u^{(n)} = Y_u^{(n)} \alpha_u$ for each $n \equiv \tilde{N}(1) = N(T)$.² Let $J_u^{(n)} = 1\{Y_u^{(n)} > 0\}$ and $b_n = M/T$.

²The superscript (n) indicates the dependence of the relevant quantity on the sample size n .

Then, following the last line above, I obtain

$$\begin{aligned}
X_1 &= \int_0^1 \frac{1}{\sqrt{b_n T}} J_u^{(n)} \dot{K} \left(\frac{v-u}{b_n} \right) \left(\frac{d\tilde{N}_u^{(n)} - Y_u^{(n)} \alpha_u du}{\sqrt{Y_v^{(n)} \alpha_v / T}} \right) \\
&= \int_0^1 \frac{1}{\sqrt{b_n}} J_u^{(n)} \sqrt{Y_u^{(n)}} \dot{K} \left(\frac{v-u}{b_n} \right) \left(\frac{d\tilde{N}_u^{(n)} / Y_u^{(n)} - \alpha_u du}{\sqrt{\alpha_v}} \right) \\
&= \sqrt{nb_n} \int_0^1 \frac{1}{b_n} J_u^{(n)} \sqrt{\frac{Y_u^{(n)}}{n}} \dot{K} \left(\frac{v-u}{b_n} \right) \left(\frac{d\tilde{N}_u^{(n)} / Y_u^{(n)} - \alpha_u du}{\sqrt{\alpha_v}} \right). \tag{A.3}
\end{aligned}$$

Theorem 4.2.2 of R-H states that if (i) $nJ^{(n)} / Y^{(n)} \rightarrow^P 1/\varsigma$ uniformly around v as $n \rightarrow \infty$; and (ii) α and ς are continuous at v , then

$$\sqrt{nb_n} \int_0^1 \frac{1}{b_n} J_u^{(n)} \dot{K} \left(\frac{v-u}{b_n} \right) \left(\frac{d\tilde{N}_u^{(n)}}{Y_u^{(n)}} - \alpha_u du \right) \rightarrow^d N \left(0, \frac{\dot{k}_2 \alpha_v}{\varsigma_v} \right) \tag{A.4}$$

as $n \rightarrow \infty$, $b_n \rightarrow 0$ and $nb_n \rightarrow \infty$. By picking $Y_u^{(n)} \equiv T$ and noting the twice continuous differentiability of $\tilde{\lambda}_u$ assumed by the theorem, assumptions (i) and (ii) are automatically satisfied. This implies that $X_1 \rightarrow^d N(0, \dot{k}_2)$ as $T \rightarrow \infty$, $M \rightarrow \infty$ and $M/T \rightarrow \infty$.

To complete the proof, it suffices to show that the second term X_2 of (A.2) is asymptotically negligible relative to the first term, which was just shown to be $O_P(1)$. Indeed, by symmetry of the kernel $\dot{K}(\cdot)$ and the twice continuous differentiability of $\tilde{\lambda}_u$, I obtain

$$\begin{aligned}
\lambda_{T_v}^* - \lambda_{T_v} &= \frac{1}{T} \int_0^1 \frac{T}{M} \dot{K} \left(\frac{v-u}{M/T} \right) \lambda_{Tu} T du - \lambda_{T_v} \\
&= \int_0^1 \frac{1}{m} \dot{K} \left(\frac{v-u}{m} \right) \tilde{\lambda}_u du - \lambda_{T_v} \\
&= \int_0^1 \dot{K}(x) \tilde{\lambda}_{v-mx} dx - \lambda_{T_v} \\
&= \tilde{\lambda}_v - \lambda_{T_v} + m \tilde{\lambda}'_v \int_0^1 x \dot{K}(x) dx + O_P(m^2) \\
&= O_P \left(\left(\frac{M}{T} \right)^2 \right).
\end{aligned}$$

If $M^5/T^4 \rightarrow 0$ (which corresponds to $nb_n^5 \rightarrow 0$), then $X_2 = \sqrt{M}(\lambda_{T_v}^* - \lambda_{T_v}) / \sqrt{\lambda_{T_v}} = O_p(M^{2.5}/T^2) = o_p(1)$, and thus is asymptotically negligible relative to X_1 .

For mean-squared consistency of $\hat{\lambda}_{T_v}$, simply apply Proposition 3.2.2 of R-H.

A.6 Proof of Theorem 7

For notational simplicity, I only treat the case where $I = [0, T]$. Under the null hypothesis, the innovations from the two processes are uncorrelated, which implies that $E(dN_s^a dN_{s+\ell}^b) = E(dN_s^a) E(dN_{s+\ell}^b) = \lambda^a \lambda^b ds d\ell$, so that

$$\begin{aligned} E(Q^s) &= \frac{1}{T^2} \int_I \int_J w_B(\ell) E(dN_s^a dN_{s+\ell}^b) \\ &= \frac{\lambda^a \lambda^b}{T^2} \int_I \int_J w_B(\ell) ds d\ell \\ &= \frac{\lambda^a \lambda^b}{T} \int_I w_B(\ell) \left(1 - \frac{|\ell|}{T}\right) d\ell. \end{aligned}$$

Before computing the variance, let us recall that the second-order reduced product density of N^k (which exists by assumption (A2)) was defined by $\varphi^{kk}(u) dt du = E(dN_t^k dN_{t+u}^k)$ for $u \neq 0$, and the unconditional autocovariance density function can thus be expressed as $c^{kk}(u) dt du = E(dN_t^k - \lambda^k dt)(dN_{t+u}^k - \lambda^k du) = \left[\varphi^{kk}(u) - (\lambda^k)^2\right] dt du$ for $u \neq 0$. Then, under the null hypothesis, I obtain

$$\begin{aligned} E((Q^s)^2) &= \frac{1}{T^4} \iint_{I^2} \iint_{J^2} w_B(\ell_1) w_B(\ell_2) E(dN_{s_1}^a dN_{s_2}^a dN_{s_1+\ell_1}^b dN_{s_2+\ell_2}^b) \\ &= \frac{1}{T^4} \iint_{I^2} \iint_{J^2} w_B(\ell_1) w_B(\ell_2) E(dN_{s_1}^a dN_{s_2}^a) E(dN_{s_1+\ell_1}^b dN_{s_2+\ell_2}^b). \end{aligned}$$

I can decompose the differential as follows:

$$\begin{aligned}
& E \left(dN_{s_1}^a dN_{s_2}^a dN_{s_1+\ell_1}^b dN_{s_2+\ell_2}^b \right) \\
&= E \left(dN_{s_1}^a dN_{s_2}^a \right) E \left(dN_{s_1+\ell_1}^b dN_{s_2+\ell_2}^b \right) \\
&= \left[E \left(dN_{s_1}^a dN_{s_2}^a \right) - (\lambda^a)^2 ds_1 ds_2 \right] \left[E \left(dN_{s_1+\ell_1}^b dN_{s_2+\ell_2}^b \right) - (\lambda^b)^2 d\ell_1 d\ell_2 \right] \\
&+ (\lambda^b)^2 \left[E \left(dN_{s_1}^a dN_{s_2}^a \right) - (\lambda^a)^2 ds_1 ds_2 \right] d\ell_1 d\ell_2 \\
&+ (\lambda^a)^2 \left[E \left(dN_{s_1+\ell_1}^b dN_{s_2+\ell_2}^b \right) - (\lambda^b)^2 d\ell_1 d\ell_2 \right] ds_1 ds_2 \\
&+ (\lambda^a)^2 (\lambda^b)^2 ds_1 ds_2 d\ell_1 d\ell_2 \\
&= c^{aa}(s_2 - s_1) c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2 \\
&+ (\lambda^b)^2 c^{aa}(s_2 - s_1) ds_1 ds_2 d\ell_1 d\ell_2 \\
&+ (\lambda^a)^2 c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2 \\
&+ (\lambda^a)^2 (\lambda^b)^2 ds_1 ds_2 d\ell_1 d\ell_2.
\end{aligned}$$

Note that the integral term associated with the last differential is $[E(Q^s)]^2$, so that

$$\begin{aligned}
Var(Q^s) &= \frac{1}{T^4} \iint_{J^2} \iint_{J^2} w_B(\ell_1) w_B(\ell_2) c^{aa}(s_2 - s_1) c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2 \\
&+ \frac{1}{T^4} \iint_{J^2} \iint_{J^2} w_B(\ell_1) w_B(\ell_2) (\lambda^b)^2 c^{aa}(s_2 - s_1) ds_1 ds_2 d\ell_1 d\ell_2 \\
&+ \frac{1}{T^4} \iint_{J^2} \iint_{J^2} w_B(\ell_1) w_B(\ell_2) (\lambda^a)^2 c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2 \\
&= A_1 + A_2 + A_3.
\end{aligned}$$

Suppose $I = [0, T]$. I evaluate the three terms individually as follows.

(i) the first term becomes

$$A_1 = \frac{2}{T^4} \int_0^T \int_0^{\ell_2} \int_{J_2} \int_{J_1} w_B(\ell_1) w_B(\ell_2) c^{aa}(s_2 - s_1) c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2.$$

where $J_i = [0, T - \ell_i]$ for $i = 1, 2$. With a change of variables

$$(s_1, s_2, \ell_1, \ell_2) \mapsto (v = s_2 - s_1, s_2, u = \ell_2 - \ell_1, \ell_2),$$

I can rewrite A_1 into

$$\begin{aligned} A_1 &= \frac{2}{T^4} \int_0^T \int_u^T \int_{-T}^{T-\ell_2} \int_0^{T-\ell_2} w_B(\ell_2 - u) w_B(\ell_2) c^{aa}(v) c^{bb}(v + u) ds_2 dv d\ell_2 du \\ &= \frac{2}{T^3} \int_0^T \int_u^T w_B(\ell_2 - u) w_B(\ell_2) \left(1 - \frac{\ell_2}{T}\right) \int_{-T}^{T-\ell_2} c^{aa}(v) c^{bb}(v + u) dv d\ell_2 du. \end{aligned}$$

To simplify further, I rely on the assumption that the bandwidth of $w(\ell) \equiv w_B(\ell)$ is small relative to T , i.e. $B = o(T)$. Then, the integral $\int_{-T}^{T-\ell_2} c^{aa}(v) c^{bb}(v + u) dv$ can be well approximated by $\Gamma(u) := \int_{-T}^T c^{aa}(v) c^{bb}(v + u) dv$, and hence

$$A_1 \approx \frac{2}{T^3} \int_0^T \mathcal{W}_2(u) \Gamma(u) du.$$

where we defined a new weighting function by $\mathcal{W}_2(u) := \int_u^T w_B(\ell - u) w_B(\ell) \left(1 - \frac{\ell}{T}\right) d\ell$.

Figure 12 gives a plot of $\mathcal{W}_2(u)$ when $w(\cdot)$ is a standard normal density function and T is large ($T \gg 3$).

(ii) With a change of variables $(s_1, s_2) \mapsto (v = s_2 - s_1, s_2)$, the second term becomes

$$\begin{aligned} A_2 &= \frac{(\lambda^b)^2}{T^4} \iint_{J^2} \iint_{J^2} w_B(\ell_1) w_B(\ell_2) \gamma^a(s_2 - s_1) ds_1 ds_2 d\ell_1 d\ell_2 \\ &= \frac{(\lambda^b)^2}{T^4} \int_0^T \int_0^T w_B(\ell_1) w_B(\ell_2) \int_{-(T-\ell_1)}^{T-\ell_2} \int_v^{v+T-\ell_1} c^{aa}(v) ds_2 dv d\ell_1 d\ell_2 \\ &= \frac{(\lambda^b)^2}{T^3} \int_0^T \int_0^T w_B(\ell_1) w_B(\ell_2) \left(1 - \frac{\ell_1}{T}\right) \int_{-(T-\ell_1)}^{T-\ell_2} c^{aa}(v) dv d\ell_1 d\ell_2. \end{aligned}$$

To simplify further, I rely on the assumption that the bandwidth B of the weighting function $w_B(\cdot)$ is small relative to T , i.e. $B = o(T)$. Then, the following holds approximately:

$$\int_{-(T-\ell_1)}^{T-\ell_2} c^{aa}(v) dv \approx \int_{-T}^T c^{aa}(v) dv.$$

As a result, we obtain

$$A_2 \approx \frac{2(\lambda^b)^2}{T^3} \omega_1 \int_0^T c^{aa}(v) dv$$

where I defined the constant $\omega_1 := \int_0^T w_B(\ell) \left(1 - \frac{\ell}{T}\right) d\ell = \int_0^{T/B} w(u) \left(1 - \frac{Bu}{T}\right) du$.

(iii) With a change of variables $(s_1, s_2) \mapsto (x = s_2 - s_1 + \ell_2 - \ell_1, s_2)$, the third term becomes

$$\begin{aligned} A_3 &= \frac{(\lambda^a)^2}{T^4} \iint_{J^2} \iint_{J^2} w_B(\ell_1) w_B(\ell_2) c^{bb}(s_2 - s_1 + \ell_2 - \ell_1) ds_1 ds_2 d\ell_1 d\ell_2 \\ &= \frac{(\lambda^a)^2}{T^4} \int_0^T \int_0^T w_B(\ell_1) w_B(\ell_2) \int_{-(T-\ell_2)}^{T-\ell_1} \int_{x+\ell_2-\ell_1}^{x+T-\ell_1+\ell_2-\ell_1} c^{bb}(x) ds_2 dx d\ell_1 d\ell_2 \\ &= \frac{(\lambda^a)^2}{T^3} \int_0^T \int_0^T w_B(\ell_1) w_B(\ell_2) \left(1 - \frac{\ell_1}{T}\right) \int_{-(T-\ell_2)}^{T-\ell_1} c^{bb}(x) dx d\ell_1 d\ell_2. \end{aligned}$$

To simplify further, I rely on the assumption that the bandwidth B of the weighting function $w_B(\cdot)$ is small relative to T , i.e. $B = o(T)$. Then, the following holds approximately:

$$\int_{-(T-\ell_2)}^{T-\ell_1} c^{bb}(v) dv \approx \int_{-T}^T c^{bb}(v) dv.$$

As a result, I obtain

$$A_3 \approx \frac{2(\lambda^a)^2}{T^3} \omega_1 \int_0^T c^{bb}(v) dv.$$

Combining the above three terms A_i for $i = 1, 2, 3$, I obtain an approximation to the variance of Q^s :

$$\text{Var}(Q^s) \approx \frac{2}{T^3} \left[\int_0^T \mathcal{W}_2(u) \Gamma(u) du + (\lambda^b)^2 \omega_1 \int_0^T c^{aa}(v) dv + (\lambda^a)^2 \omega_1 \int_0^T c^{bb}(v) dv \right].$$

A.7 Proof of (2.21)

For notational convenience, I drop the superscript k from all relevant symbols throughout this proof. Let $R/T \rightarrow 0$ as $T \rightarrow \infty$. I start by decomposing $\hat{c}_R(\ell) \equiv$

$\hat{c}_{R^k}^{kk}(\ell)$ as follows:

$$\begin{aligned}
\hat{c}_R(\ell) &= \frac{1}{TR} \int_0^T \int_0^T \ddot{K}\left(\frac{t-s-\ell}{R}\right) \left(dN_s - \frac{N_T}{T} ds\right) \left(dN_t - \frac{N_T}{T} dt\right) \\
&= \frac{1}{TR} \int_0^T \int_0^T \ddot{K}\left(\frac{t-s-\ell}{R}\right) dN_s dN_t - \frac{1}{TR} \frac{N_T}{T} \int_0^T \int_0^T \ddot{K}\left(\frac{t-s-\ell}{R}\right) ds dN_t \\
&\quad - \frac{1}{TR} \frac{N_T}{T} \int_0^T \int_0^T \ddot{K}\left(\frac{t-s-\ell}{R}\right) dN_s dt + \frac{1}{TR} \frac{N_T}{T} \frac{N_T}{T} \int_0^T \int_0^T \ddot{K}\left(\frac{t-s-\ell}{R}\right) ds dt \\
&=: C_1 + C_2 + C_3 + C_4.
\end{aligned}$$

Now, the second term is

$$\begin{aligned}
C_2 &= -\frac{1}{TR} \frac{N_T}{T} \int_0^T \int_0^T \ddot{K}\left(\frac{t-s-\ell}{R}\right) ds dN_t^b \\
&= -\frac{N_T}{T^2} \sum_{j=1}^{N_T} \int_0^T \frac{1}{R} \ddot{K}\left(\frac{t_j-s-\ell}{R}\right) ds = -\frac{N_T}{T^2} \sum_{j=1}^{N_T} \int_{(t_j^b-T-\ell)/R}^{(t_j^b-\ell)/R} \ddot{K}(x) dx \\
&= -\frac{N_T}{T^2} \sum_{j=1}^{N_T} \left[1_{\{\ell < t_j < T+\ell\} \cap [0, T]} + o(1) \right] \\
&= -\frac{N_T}{T^2} [N_{T \wedge (T+\ell)} - N_{\ell \vee 0}] + o(1),
\end{aligned}$$

where the third equality made use of assumption (A4c). By stationarity of N^k , I observe that $N_{T \wedge (T+\ell)} - N_{\ell \vee 0} = \frac{T-|\ell|}{T} N_T$. Therefore, up to the leading term,

$$C_2 = -\frac{N_T^2}{T^2} \left(1 - \frac{|\ell|}{T}\right).$$

Similarly, by stationarity of N^b , the third term is, up to the leading term,

$$C_3 = -\frac{N_T^2}{T^2} \left(1 - \frac{|\ell|}{T}\right) = C_2.$$

The last term is

$$\begin{aligned}
C_4 &= \frac{1}{T} \frac{N_T}{T} \frac{N_T}{T} \int_0^T \int_0^T \frac{1}{R} \ddot{K}\left(\frac{t-s-\ell}{R}\right) ds dt \\
&= \frac{N_T^2}{T^2} \frac{1}{T} \int_0^T \int_{(t-T-\ell)/R}^{(t-\ell)/R} \ddot{K}(x) dx dt \\
&= \frac{N_T^2}{T^2} \frac{1}{T} \int_0^T \left[1_{\{0 < t-\ell < T\} \cap [0, T]} + o(1) \right] dt \\
&= \frac{N_T^2}{T^2} \left(1 - \frac{|\ell|}{T}\right) + o(1),
\end{aligned}$$

which is $-C_2$ (neglecting the $o(1)$ terms). As a result, except for the $o(1)$ terms, I obtain

$$\hat{c}_R(\ell) = \frac{1}{TR} \sum_{i=1}^{N_T} \sum_{j=1}^{N_T} \ddot{K}\left(\frac{t_j - t_i - \ell}{R}\right) - \frac{N_T^2}{T^2} \left(1 - \frac{|\ell|}{T}\right),$$

which is (2.21).

A.8 Proof of Theorem 8

Let $d\hat{\epsilon}_t^k = dN_t^k - \hat{\lambda}_t^k dt$ for $k = a, b$. Then,

$$\begin{aligned} Q &= \int_I w_B(\ell) \hat{\gamma}_H^2(\ell) d\ell \\ &= \int_I w_B(\ell) \frac{1}{(TH)^2} \iiint \int_{(0,T]^4} K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b d\ell \\ &= \frac{1}{T^2} \iiint \int_{(0,T]^4} \int_I w(\ell) \frac{1}{H^2} K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b \end{aligned}$$

A.8.1 Asymptotic Mean of Q

By Fubini's theorem, the expectation of Q becomes an multiple integration with respect to $E[d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{s_1+u}^b d\hat{\epsilon}_{s_2+u}^b]$, which, under the null hypothesis (2.11), can be split into $E[d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a] E[d\hat{\epsilon}_{s_1+u}^b d\hat{\epsilon}_{s_2+u}^b]$. By the law of iterated expectations and the martingale property of the innovations $d\hat{\epsilon}_u^k$, it follows that $E[d\hat{\epsilon}_{u_1}^k d\hat{\epsilon}_{u_2}^k] = 0$ unless $u_1 = u_2 = u$ when it is equal to $E[(d\hat{\epsilon}_u^k)^2]$. Then, I can simplify the differential

$(d\hat{\epsilon}_u^k)^2$ as follows:

$$\begin{aligned}
(d\hat{\epsilon}_u^k)^2 &= (dN_u^k - \hat{\lambda}_u^k du)^2 \\
&= (dN_u^k - \lambda_u^k du + \lambda_u^k du - \hat{\lambda}_u^k du)^2 \\
&= (dN_u^k - \lambda_u^k du)^2 + (\lambda_u^k du - \hat{\lambda}_u^k du)^2 + 2(dN_u^k - \lambda_u^k du)(\lambda_u^k du - \hat{\lambda}_u^k du) \\
&= (dN_u^k - \lambda_u^k du)^2 + o_P(du) \\
&= dN_u^k - 2dN_u^k \lambda_u^k du + (\lambda_u^k du)^2 + o_P(du) \\
&= dN_u^k + o_P(du).
\end{aligned} \tag{A.5}$$

The second-to-last equality holds because of assumption (A1), which implies that $(dN_u^k)^2 = dN_u^k$ almost surely; hence the second order differential $(d\hat{\epsilon}_u^k)^2$ has a dominating first-order increment dN_u^k . It is therefore true, up to $O_P(du)$, that $E[(d\hat{\epsilon}_u^k)^2] = E[dN_u^k] = \lambda^k du$.

Now, letting $b = B/T$ and $h = H/T$, the expected value of Q is evaluated as follows:

$$\begin{aligned}
E(Q) &= \frac{1}{T^2 H^2} \iint_{(0,T]^2} \int_I w_B(\ell) K^2\left(\frac{t-s-\ell}{H}\right) d\ell \lambda^a \lambda^b ds dt \\
&= \frac{T^3}{T^2 H^2} \iint_{(0,1]^2} \int_{I/T} w_B(T\sigma) K^2\left(\frac{v-u-\sigma}{H/T}\right) d\sigma \lambda^a \lambda^b du dv \\
&= \frac{1}{T^2 h^2} \iint_{(0,1]^2} \int_{I/T} w_b(\sigma) K^2\left(\frac{v-u-\sigma}{h}\right) d\sigma \lambda^a \lambda^b du dv
\end{aligned}$$

Then, as $h \rightarrow 0$,

$$\begin{aligned}
&\iint_{(0,1]^2} \frac{1}{h^2} K^2\left(\frac{v-u-\sigma}{h}\right) du dv \\
&= \frac{1}{h} \int_0^1 \int_{(v-1-\sigma)/h}^{(v-\sigma)/h} K^2(x) dx dv \\
&= \frac{1}{h} \int_0^1 1_{\{0 < v-\sigma < 1\} \cap [0,1]} dv \int_{-\infty}^{\infty} K^2(x) dx + o\left(\frac{1}{h}\right) \\
&= \frac{1}{h} (1 - |\sigma|) \kappa_2 + o\left(\frac{1}{h}\right).
\end{aligned}$$

where $\kappa_2 = \int_{-\infty}^{\infty} K^2(x)dx$ (from assumption (A4a)). As a result, as $T \rightarrow \infty$, $Th = H \rightarrow \infty$ and $h = H/T \rightarrow 0$, the asymptotic mean of Q under the null hypothesis is given by

$$\begin{aligned} E(Q) &= \frac{1}{T^2 h} \lambda^a \lambda^b \kappa_2 \int_{I/T} w_b(\sigma) (1 - |\sigma|) d\sigma + o\left(\frac{1}{T^2 h}\right) \\ &= \frac{1}{T^2 h} \lambda^a \lambda^b \kappa_2 \int_I \frac{1}{b} w\left(\frac{\ell}{bT}\right) \left(1 - \frac{|\ell|}{T}\right) \frac{1}{T} d\ell + o\left(\frac{1}{T^2 h}\right) \\ &= \frac{1}{TH} \lambda^a \lambda^b \kappa_2 \int_I w_B(\ell) \left(1 - \frac{|\ell|}{T}\right) d\ell + o\left(\frac{1}{TH}\right). \end{aligned} \quad (\text{A.6})$$

From (A.5), I also observe that $(d\hat{\epsilon}_u^k)^2 = (d\epsilon_u^k)^2 + o_P(du)$, which entails that $E(Q) = E(\tilde{Q})$.

A.8.2 Asymptotic Variance of Q Under the Null

The Case Without Autocorrelations

Now, I derive the asymptotic variance of Q as $T \rightarrow \infty$, and $H/T \rightarrow 0$ as $H \rightarrow \infty$.

Let $I \equiv [c_1, c_2] \subseteq [-T, T]$, where $c_1 < c_2$. Consider

$$\begin{aligned} E(Q^2) &= \frac{1}{(TH)^4} \iint_{I^2} w(\ell_1) w(\ell_2) \int \cdots \int_{(0,T]^8} K\left(\frac{t_{11} - s_{11} - \ell_1}{H}\right) K\left(\frac{t_{21} - s_{21} - \ell_1}{H}\right) \\ &\quad K\left(\frac{t_{12} - s_{12} - \ell_2}{H}\right) K\left(\frac{t_{22} - s_{22} - \ell_2}{H}\right) E\left[d\hat{\epsilon}_{s_{11}}^a d\hat{\epsilon}_{s_{12}}^a d\hat{\epsilon}_{t_{11}}^b d\hat{\epsilon}_{t_{12}}^b d\hat{\epsilon}_{s_{21}}^a d\hat{\epsilon}_{s_{22}}^a d\hat{\epsilon}_{t_{21}}^b d\hat{\epsilon}_{t_{22}}^b\right] d\ell_1 d\ell_2. \end{aligned}$$

Assume that (i) there is no cross-correlation between the two innovation processes, i.e. $\gamma(u) = 0$; and (ii) there is no auto-correlation for each component process, i.e. $c^{aa}(u) = c^{bb}(u) = 0$. I will relax the second assumption in the next subsection.

A key observation is that $E(Q^2) \neq 0$ only in the following cases (in all cases $s_1 \neq s_2 \neq t_1 \neq t_2$ and $s \neq t$):

1. $\mathcal{R}_1 = \{s_{11} = s_{12} = s_1, s_{21} = s_{22} = s_2, t_{11} = t_{12} = t_1, t_{21} = t_{22} = t_2\};$
2. $\mathcal{R}_2 = \{s_{11} = s_{12} = s_1, s_{21} = s_{22} = s_2, t_{11} = t_{21} = t_1, t_{12} = t_{22} = t_2\};$
3. $\mathcal{R}_3 = \{s_{11} = s_{12} = s_1, s_{21} = s_{22} = s_2, t_{11} = t_{22} = t_1, t_{12} = t_{21} = t_2\};$
4. $\mathcal{R}_4 = \{s_{11} = s_{21} = s_1, s_{12} = s_{22} = s_2, t_{11} = t_{12} = t_1, t_{21} = t_{22} = t_2\};$
5. $\mathcal{R}_5 = \{s_{11} = s_{21} = s_1, s_{12} = s_{22} = s_2, t_{11} = t_{21} = t_1, t_{12} = t_{22} = t_2\};$
6. $\mathcal{R}_6 = \{s_{11} = s_{21} = s_1, s_{12} = s_{22} = s_2, t_{11} = t_{22} = t_1, t_{12} = t_{21} = t_2\};$
7. $\mathcal{R}_7 = \{s_{11} = s_{22} = s_1, s_{12} = s_{21} = s_2, t_{11} = t_{12} = t_1, t_{21} = t_{22} = t_2\};$
8. $\mathcal{R}_8 = \{s_{11} = s_{22} = s_1, s_{12} = s_{21} = s_2, t_{11} = t_{21} = t_1, t_{12} = t_{22} = t_2\};$
9. $\mathcal{R}_9 = \{s_{11} = s_{22} = s_1, s_{12} = s_{21} = s_2, t_{11} = t_{22} = t_1, t_{12} = t_{21} = t_2\};$
10. $\mathcal{R}_{10} = \{s_{11} = s_{12} = s_{21} = s_{22} = s \text{ and } t_{11} = t_{12} = t_{21} = t_{22} = t\}.$

Under the null of no cross-correlation, for cases 1 to 9, we have, up to $O(ds_1 ds_2 dt_1 dt_2)$,

$$\begin{aligned}
& E \left[d\hat{\epsilon}_{s_{11}}^a d\hat{\epsilon}_{s_{12}}^a d\hat{\epsilon}_{t_{11}}^b d\hat{\epsilon}_{t_{12}}^b d\hat{\epsilon}_{s_{21}}^a d\hat{\epsilon}_{s_{22}}^a d\hat{\epsilon}_{t_{21}}^b d\hat{\epsilon}_{t_{22}}^b \right] \\
&= E \left[\left(d\hat{\epsilon}_{s_1}^a \right)^2 \left(d\hat{\epsilon}_{s_2}^a \right)^2 \left(d\hat{\epsilon}_{t_1}^b \right)^2 \left(d\hat{\epsilon}_{t_2}^b \right)^2 \right] \\
&= E \left[\left(d\hat{\epsilon}_{s_1}^a \right)^2 \left(d\hat{\epsilon}_{s_2}^a \right)^2 \right] E \left[\left(d\hat{\epsilon}_{t_1}^b \right)^2 \left(d\hat{\epsilon}_{t_2}^b \right)^2 \right] \\
&= E \left[dN_{s_1}^a dN_{s_2}^a \right] E \left[dN_{t_1}^b dN_{t_2}^b \right] \\
&= \left\{ (\lambda^a)^2 ds_1 ds_2 + E \left[\left(dN_{s_1}^a - \lambda^a ds_1 \right) \left(dN_{s_2}^a - \lambda^a ds_2 \right) \right] \right\} \\
&\quad \left\{ (\lambda^b)^2 dt_1 dt_2 + E \left[\left(dN_{t_1}^b - \lambda^b dt_1 \right) \left(dN_{t_2}^b - \lambda^b dt_2 \right) \right] \right\} \\
&= \left[(\lambda^a)^2 + c^{aa}(s_2 - s_1) \right] \left[(\lambda^b)^2 + c^{bb}(t_2 - t_1) \right] ds_1 ds_2 dt_1 dt_2;
\end{aligned}$$

while for case 10, I have, up to $O(dsdt)$,

$$\begin{aligned}
& E \left[d\hat{\epsilon}_{s_{11}}^a d\hat{\epsilon}_{s_{12}}^a d\hat{\epsilon}_{t_{11}}^b d\hat{\epsilon}_{t_{12}}^b d\hat{\epsilon}_{s_{21}}^a d\hat{\epsilon}_{s_{22}}^a d\hat{\epsilon}_{t_{21}}^b d\hat{\epsilon}_{t_{22}}^b \right] \\
&= E \left[(d\hat{\epsilon}_s^a)^4 (d\hat{\epsilon}_t^b)^4 \right] \\
&= E \left[dN_s^a \right] E \left[dN_t^b \right] \\
&= \lambda^a \lambda^b dsdt.
\end{aligned}$$

Cases 1 and 9: the innermost eight inner integrals reduce to four integrals, so that

$$\begin{aligned}
& E(Q^2) \\
&= \frac{(\lambda^a \lambda^b)^2}{(TH)^4} \iint_{I^2} w_B(\ell_1) w_B(\ell_2) \iiint_{(0,T]^4} K\left(\frac{t_1-s_1-\ell_1}{H}\right) K\left(\frac{t_2-s_2-\ell_1}{H}\right) \\
&\quad K\left(\frac{t_1-s_1-\ell_2}{H}\right) K\left(\frac{t_2-s_2-\ell_2}{H}\right) ds_1 ds_2 dt_1 dt_2 d\ell_1 d\ell_2 \\
&= \frac{(\lambda^a \lambda^b)^2 T^4}{(TH)^4} \iint_{(I/T)^2} \frac{T^2}{B^2} w\left(\frac{\sigma_1}{B/T}\right) w\left(\frac{\sigma_2}{B/T}\right) \iint_{(0,1]^2} \int_{v_2-1}^{v_2} \int_{v_1-1}^{v_1} K\left(\frac{u-\sigma_1}{H/T}\right) K\left(\frac{v-\sigma_1}{H/T}\right) \\
&\quad K\left(\frac{u-\sigma_2}{H/T}\right) K\left(\frac{v-\sigma_2}{H/T}\right) dudvdv_1 dv_2 d\sigma_1 d\sigma_2 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \int_{c_1/T}^{c_2/T} w_b(\sigma_1) w_b(\sigma_1 - zh) \\
&\quad \iint_{(0,1]^2} \int_{\frac{\sigma_1-c_2/T}{h}}^{\frac{\sigma_1-c_1/T}{h}} \int_{\frac{v_2-1-\sigma_1}{h}}^{\frac{v_2-\sigma_1}{h}} \int_{\frac{v_1-1-\sigma_1}{h}}^{\frac{v_1-\sigma_1}{h}} K(x) K(y) K(x+z) K(y+z) dx dy dz dv_1 dv_2 d\sigma_1 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \int_{c_1/T}^{c_2/T} w_b(\sigma_1) [w_b(\sigma_1) - zh w'_b(\tilde{\sigma}_1)] \iint_{(0,1]^2} 1_{\{(v_1, v_2): 0 \vee \sigma_1 < v_i < (1+\sigma_1) \wedge 1\}} dv_1 dv_2 d\sigma_1 \\
&\quad \iiint_{\mathbb{R}^3} K(x) K(y) K(x+z) K(y+z) dx dy dz + o\left(\frac{1}{T^3 h}\right) \\
&= \frac{(\lambda^a \lambda^b)^2}{T^2 H} \kappa_4 \int_I w_B^2(\ell_1) \left(1 - \frac{|\ell_1|}{T}\right)^2 d\ell_1 + o\left(\frac{1}{T^2 H}\right)
\end{aligned}$$

I applied the change of variables: $(s_1, s_2, t_1, t_2) \mapsto (u = \frac{t_1-s_1}{T}, v = \frac{t_2-s_2}{T}, v_1 = \frac{t_1}{T}, v_2 = \frac{t_2}{T})$ in the second equality, and $(u, v, \ell_2) \mapsto (x = \frac{u-\sigma_1}{h}, y = \frac{v-\sigma_1}{h}, z = \frac{\sigma_1-\sigma_2}{h})$ in the

third equality. To get the fourth equality, I did a first-order Taylor expansion of $w_b(\sigma_1 - zh)$ around σ_1 , with $\ddot{\sigma}_1 \in [\sigma_1 - zh, \sigma_1]$.

Cases 2, 4, 6 and 8:

$$\begin{aligned}
& E(Q^2) \\
&= \frac{(\lambda^a \lambda^b)^2}{(TH)^4} \iint_{I^2} w_B(\ell_1) w_B(\ell_2) \int \cdots \int_{(0,T]^4} K\left(\frac{t_1 - s_1 - \ell_1}{H}\right) K\left(\frac{t_1 - s_2 - \ell_1}{H}\right) \\
&\quad K\left(\frac{t_2 - s_1 - \ell_2}{H}\right) K\left(\frac{t_2 - s_2 - \ell_2}{H}\right) ds_1 ds_2 dt_1 dt_2 d\ell_1 d\ell_2 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \iiint_{(0,1]^3} \int_{\frac{u_1-1}{h}}^{\frac{u_1}{h}} \int_{\frac{v_2-u_1-c_2/T}{h}}^{\frac{v_2-u_1-c_1/T}{h}} \int_{\frac{v_1-u_1-c_2/T}{h}}^{\frac{v_1-u_1-c_1/T}{h}} w_b(v_1 - u_1 - xh) w_b(v_2 - u_1 - yh) \\
&\quad K(x) K(x+z) K(y) K(y+z) dx dy dz du_1 dv_1 dv_2 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \kappa_4 \iiint_{(0,1]^3} w_b(v_1 - u_1) w_b(v_2 - u_1) 1_{\{v_1 - u_1 \in I/T\}} 1_{\{v_2 - u_1 \in I/T\}} du_1 dv_1 dv_2 + o\left(\frac{1}{T^4 h}\right) \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \kappa_4 \int_0^1 \int_{-s}^{1-s} \int_{-s}^{1-s} w_b(u) w_b(v) 1_{\{u \in I/T\}} 1_{\{v \in I/T\}} du dv ds + o\left(\frac{1}{T^4 h}\right) \\
&= O\left(\frac{1}{T^3 H}\right).
\end{aligned}$$

I applied the change of variables: $(\ell_1, \ell_2, s_2) \mapsto (x = \frac{v_1 - u_1 - \sigma_1}{H/T}, y = \frac{v_2 - u_1 - \sigma_2}{H/T}, z = \frac{u_1 - u_2}{H/T})$ in the second equality, and $(u_1, v_1, v_2) \mapsto (s, u = v_1 - u_1, v = v_2 - u_1)$ in the fourth equality, and the fact that $\int_0^1 \int_{-s}^{1-s} \int_{-s}^{1-s} w_b(u) w_b(v) du dv ds = O(1)$ in the last equality.

Cases 3 and 7:

$$\begin{aligned}
& E(Q^2) \\
&= \frac{(\lambda^a \lambda^b)^2}{(TH)^4} \iint_{I^2} w_B(\ell_1) w_B(\ell_2) \int \cdots \int_{(0,T]^4} K\left(\frac{t_1 - s_1 - \ell_1}{H}\right) K\left(\frac{t_2 - s_1 - \ell_1}{H}\right) \\
&\quad K\left(\frac{t_1 - s_2 - \ell_2}{H}\right) K\left(\frac{t_2 - s_2 - \ell_2}{H}\right) ds_1 ds_2 dt_1 dt_2 d\ell_1 d\ell_2 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \iiint_{(0,1]^3} \int_{\frac{v_1 - u_2 - c_1/T}{h}}^{\frac{v_1 - u_2 - c_1/T}{h}} \int_{\frac{v_2 - \sigma_1}{h}}^{\frac{v_2 - \sigma_1}{h}} \int_{\frac{v_1 - u_1 - c_1/T}{h}}^{\frac{v_1 - u_1 - c_1/T}{h}} w_b(v_1 - u_1 - xh) w_b(v_1 - u_2 - yh) \\
&\quad K(x) K(x+z) K(y) K(y+z) dx dy dz du_1 du_2 dv_1 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \kappa_4 \iiint_{(0,1]^3} w_b(v_1 - u_1) w_b(v_1 - u_2) 1_{\{v_1 - u_1 \in I/T\}} 1_{\{v_1 - u_2 \in I/T\}} du_1 du_2 dv_1 + o\left(\frac{1}{T^4 h}\right) \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h} \kappa_4 \int_0^1 \int_{t-1}^t \int_{t-1}^t w_b(u) w_b(v) 1_{\{u \in I/T\}} 1_{\{v \in I/T\}} du dv dt + o\left(\frac{1}{T^4 h}\right) \\
&= O\left(\frac{1}{T^3 H}\right).
\end{aligned}$$

I applied the change of variables: $(\ell_1, \ell_2, s_2) \mapsto (x = \frac{v_1 - u_1 - \sigma_1}{H/T}, y = \frac{v_1 - u_2 - \sigma_2}{H/T}, z = \frac{v_2 - v_1}{H/T})$ in the second equality, and $(v_1, u_1, u_2) \mapsto (t, u = v_1 - u_1, v = v_1 - u_2)$ in the fourth equality, and the fact that $\int_0^1 \int_{t-1}^t \int_{t-1}^t w_b(u) w_b(v) du dv dt = O(1)$ in the last equality.

Case 5:

$$\begin{aligned}
& E(Q^2) \\
&= \frac{(\lambda^a \lambda^b)^2}{(TH)^4} \iint_{I^2} w_B(\ell_1) w_B(\ell_2) \int \cdots \int_{(0,T]^4} K^2\left(\frac{t_1-s_1-\ell_1}{H}\right) K^2\left(\frac{t_2-s_2-\ell_2}{H}\right) ds_1 ds_2 dt_1 dt_2 d\ell_1 d\ell_2 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h^2} \iiint_{(0,1]^3} \int_{\frac{v_2-u_1-c_1}{h}}^{\frac{v_2-u_1-c_1}{h}+T} \int_{\frac{v_1-u_1-c_1}{h}}^{\frac{v_1-u_1-c_1}{h}+T} w_b(v_1-u_1-xh) w_b(v_2-u_2-yh) \\
& \quad K^2(x) K^2(y) dx dy du_1 du_2 dv_1 dv_2 \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h^2} \kappa_2^2 \iiint_{(0,1]^3} w_b(v_1-u_1) w_b(v_2-u_2) 1_{\{v_1-u_1 \in I/T\}} 1_{\{v_2-u_2 \in I/T\}} du_1 du_2 dv_1 dv_2 + o\left(\frac{1}{T^2 h^2}\right) \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h^2} \kappa_2^2 \int_0^1 \int_0^1 \int_{-u_2}^{1-u_2} \int_{-u_1}^{1-u_1} w_b(u) w_b(v) 1_{\{u \in I/T\}} 1_{\{v \in I/T\}} du dv du_1 du_2 + o\left(\frac{1}{T^2 h^2}\right) \\
&= \frac{(\lambda^a \lambda^b)^2}{T^4 h^2} \kappa_2^2 \left(\int_0^1 \int_{-s}^{1-s} w_b(u) 1_{\{u \in I/T\}} du ds \right)^2 + o\left(\frac{1}{T^4 h^2}\right) \\
&= [E(Q)]^2 + o\left(\frac{1}{T^2 H^2}\right).
\end{aligned}$$

I applied the change of variables: $(\ell_1, \ell_2) \mapsto (x = \frac{v_1-u_1-\sigma_1}{H/T}, y = \frac{v_2-u_2-\sigma_2}{H/T})$ in the second equality, and $(u_1, u_2, v_1, v_2) \mapsto (u_1, u_2, u = v_1 - u_1, v = v_2 - u_2)$ in the fourth equality. The last equality follows from Fubini's theorem, which gives

$$\begin{aligned}
\int_0^1 \int_{-s}^{1-s} w_b(u) 1_{\{u \in I/T\}} du ds &= \left(\int_{-1}^0 \int_{-u}^1 + \int_0^1 \int_0^{1-u} \right) [w_b(u) 1_{\{u \in I/T\}}] ds du \\
&= \int_{c_1/T}^{c_2/T} (1-|u|) w_b(u) du. \\
&= \int_{c_1}^{c_2} \left(1 - \frac{|\ell|}{T}\right) w_B(\ell) d\ell = E(Q)
\end{aligned}$$

Case 10:

$$\begin{aligned}
E(Q^2) &= \frac{\lambda^a \lambda^b}{(TH)^4} \iint_{I^2} w_B(\ell_1) w_B(\ell_2) \iint_{(0,T]^2} K^2\left(\frac{t-s-\ell_1}{H}\right) K^2\left(\frac{t-s-\ell_2}{H}\right) ds dt d\ell_1 d\ell_2 \\
&= \frac{\lambda^a \lambda^b}{T^4 h^2} \iint_{(0,1]^2} \int_{\frac{v-u-c_2/T}{h}}^{\frac{v-u-c_1/T}{h}} \int_{\frac{v-u-c_2/T}{h}}^{\frac{v-u-c_1/T}{h}} w_b(v-u-xh) w_b(v-u-yh) K^2(x) K^2(y) dx dy du dv \\
&= \frac{\lambda^a \lambda^b}{T^4 h^2} \kappa_2^2 \int_0^1 \int_{-u}^{1-u} w_b^2(r) 1_{\{r \in I/T\}} dr du + o\left(\frac{1}{T^5 h^2}\right) \\
&= \frac{\lambda^a \lambda^b}{T^4 h^2} \kappa_2^2 \int_{I/T} (1-|r|) w_b^2(r) dr + o\left(\frac{1}{T^5 h^2}\right) \\
&= \frac{\lambda^a \lambda^b}{T^5 h^2} \kappa_2^2 \int_I \left(1 - \frac{|\ell|}{T}\right) w_B^2(\ell) d\ell + o\left(\frac{1}{T^5 h^2}\right) \\
&= O\left(\frac{1}{T^3 H^2}\right).
\end{aligned}$$

I applied the change of variables: $(\ell_1, \ell_2) \mapsto (x = \frac{v-u-\sigma_1}{H/T}, y = \frac{v-u-\sigma_2}{H/T})$ in the second equality, and $(u, v) \mapsto (u, r = v - u)$ in the third equality. The second-to-last equality follows from Fubini's theorem.

We observe that the leading terms of the asymptotic variance come from cases 1 and 9 only, thus we conclude that, as $T \rightarrow \infty$ and $H/T \rightarrow 0$ as $H \rightarrow \infty$,

$$\begin{aligned}
\text{Var}(Q) &= E(Q^2) - [E(Q)]^2 \\
&= 2 \frac{(\lambda^a \lambda^b)^2}{T^2 H} \kappa_4 \int_I w_B^2(\ell_2) \left(1 - \frac{|\ell_2|}{T}\right)^2 d\ell_2 + o\left(\frac{1}{T^2 H}\right). \tag{A.7}
\end{aligned}$$

The Case With Autocorrelations

Suppose the two point processes N^a and N^b exhibit autocorrelations, i.e. $c^{aa}(u)$ and $c^{bb}(u)$ are not identically zero. Then, it is necessary to modify the asymptotic

variance of Q . I start by noting that, up to $O(ds_1 ds_2 dt_1 dt_2)$,

$$\begin{aligned}
& E \left[\left(d\hat{\epsilon}_{s_1}^a \right)^2 \left(d\hat{\epsilon}_{s_2}^a \right)^2 \right] E \left[\left(d\hat{\epsilon}_{t_1}^b \right)^2 \left(d\hat{\epsilon}_{t_2}^b \right)^2 \right] \\
&= E \left[dN_{s_1}^a dN_{s_2}^a \right] E \left[dN_{t_1}^b dN_{t_2}^b \right] \\
&= \left\{ (\lambda^a)^2 ds_1 ds_2 + E \left[\left(dN_{s_1}^a - \lambda^a ds_1 \right) \left(dN_{s_2}^a - \lambda^a ds_2 \right) \right] \right\} \\
&\quad \left\{ (\lambda^b)^2 dt_1 dt_2 + E \left[\left(dN_{t_1}^b - \lambda^b dt_1 \right) \left(dN_{t_2}^b - \lambda^b dt_2 \right) \right] \right\} \\
&= \left[(\lambda^a)^2 + c^{aa}(s_2 - s_1) \right] \left[(\lambda^b)^2 + c^{bb}(t_2 - t_1) \right] ds_1 ds_2 dt_1 dt_2.
\end{aligned}$$

As before, I split the computation into 10 separate cases. Since the computation techniques are analogous to the case without autocorrelations, let's focus on cases 1 and 9 which yield the dominating terms for $Var(Q)$. Under cases 1 and 9:

$$\begin{aligned}
& E(Q^2) \\
&= \frac{1}{(TH)^4} \iint_{I^2} w_B(\ell_1) w_B(\ell_2) \iiint_{(0,T]^4} K\left(\frac{t_1-s_1-\ell_1}{H}\right) K\left(\frac{t_2-s_2-\ell_1}{H}\right) K\left(\frac{t_1-s_1-\ell_2}{H}\right) K\left(\frac{t_2-s_2-\ell_2}{H}\right) \\
&\quad [(\lambda^a)^2 + c^{aa}(s_2 - s_1)] \left[(\lambda^b)^2 + c^{bb}(t_2 - t_1) \right] ds_1 ds_2 dt_1 dt_2 d\ell_1 d\ell_2 \\
&= \frac{T^4}{(TH)^4} \iint_{(I/T)^2} \frac{T^2}{B^2} w\left(\frac{\sigma_1}{B/T}\right) w\left(\frac{\sigma_2}{B/T}\right) \iint_{(0,1]^2} \int_{v_2-1}^{v_2} \int_{v_1-1}^{v_1} K\left(\frac{u-\sigma_1}{H/T}\right) K\left(\frac{v-\sigma_1}{H/T}\right) K\left(\frac{u-\sigma_2}{H/T}\right) K\left(\frac{v-\sigma_2}{H/T}\right) \\
&\quad [(\lambda^a)^2 + c^{aa}(T(u - v + v_2 - v_1))] \left[(\lambda^b)^2 + c^{bb}(T(v_2 - v_1)) \right] dudvdv_1 dv_2 d\sigma_1 d\sigma_2 \\
&= \frac{1}{T^4 h} \int_{c_1/T}^{c_2/T} w_b(\sigma_1) w_b(\sigma_1 - zh) \iint_{(0,1]^2} \int_{\frac{\sigma_1-c_2/T}{h}}^{\frac{\sigma_1-c_1/T}{h}} \int_{\frac{v_2-1-\sigma_1}{h}}^{\frac{v_2-\sigma_1}{h}} \int_{\frac{v_1-1-\sigma_1}{h}}^{\frac{v_1-\sigma_1}{h}} K(x) K(y) K(x+z) K(y+z) \\
&\quad [(\lambda^a)^2 + c^{aa}(Th(x-y) + T(v_2 - v_1))] \left[(\lambda^b)^2 + c^{bb}(T(v_2 - v_1)) \right] dx dy dz dv_1 dv_2 d\sigma_1 \\
&= \frac{1}{T^4 h} \int_{c_1/T}^{c_2/T} w_b^2(\sigma_1) \iiint_{\mathbb{R}^3} K(x) K(y) K(x+z) K(y+z) dx dy dz \iint_{(0,1]^2} 1_{\{(v_1, v_2): 0 \vee \sigma_1 < v_i < (1+\sigma_1) \wedge 1\}} \\
&\quad [(\lambda^a)^2 + c^{aa}(T(v_2 - v_1))] \left[(\lambda^b)^2 + c^{bb}(T(v_2 - v_1)) \right] dv_1 dv_2 d\sigma_1 + o\left(\frac{1}{T^3 h}\right) \\
&= \frac{1}{T^2 H} \kappa_4 \int_I w_B^2(\ell_1) \int_{-(T-|\ell_1|)}^{T-|\ell_1|} \left(1 - \frac{|r|}{T} - \frac{|\ell_1|}{T}\right) [(\lambda^a)^2 + c^{aa}(r)] \left[(\lambda^b)^2 + c^{aa}(r) \right] dr d\ell_1 + o\left(\frac{1}{T^2 H}\right).
\end{aligned}$$

I applied the change of variables: $(s_1, s_2, t_1, t_2) \mapsto (u = \frac{t_1-s_1}{T}, v = \frac{t_2-s_2}{T}, v_1 = \frac{t_1}{T}, v_2 = \frac{t_2}{T})$, $b = B/T$ and $h = H/T$ in the second equality, and $(u, v, \ell_2) \mapsto (x = \frac{u-\sigma_1}{h}, y = \frac{v-\sigma_1}{h}, z = \frac{\sigma_1-\sigma_2}{h})$ in the third equality. To get the fourth equality, I did a first-order Taylor expansion of $w_b(\sigma_1 - zh)$ around σ_1 , with $\sigma_1 \in [\sigma_1 - zh, \sigma_1]$.

To get the last equality, I let $g(x) = [(\lambda^a)^2 + c^{aa}(Tx)] \left[(\lambda^b)^2 + c^{bb}(Tx) \right]$, and let $r = v_2 - v_1$. Suppose $\sigma_1 > 0$. Then, by Fubini's theorem, the innermost double integration (with respect to v_1 and v_2) becomes

$$\begin{aligned}
& \int_{\sigma_1}^1 \int_{-v_1}^{1-v_1} g(r) dr dv_1 \\
&= \left(\int_{-(1-\sigma_1)}^{-\sigma_1} \int_{-r}^1 + \int_{-\sigma_1}^0 \int_{\sigma_1}^1 + \int_0^{1-\sigma_1} \int_{\sigma_1}^{1-r} \right) g(r) dv_1 dr \\
&= \int_{-1}^{-\sigma_1} (1+r) g(r) dr - \int_{-1}^{-(1-\sigma_1)} (1+r) g(r) dr + \int_{-\sigma_1}^0 (1-\sigma_1) g(r) dr \\
&+ \int_0^{1-\sigma_1} (1-r-\sigma_1) g(r) dr.
\end{aligned}$$

The first integral can be simplified to

$$\begin{aligned}
\int_{-1}^{-\sigma_1} (1+r) g(r) dr &= \int_{-(1-\sigma_1)}^0 (1+r-\sigma_1) g(r-\sigma_1) dr \\
&= \int_{-(1-\sigma_1)}^0 (1+r-\sigma_1) [g(r) - \dot{\sigma}_1 g'(r)] dr
\end{aligned}$$

for some $\dot{\sigma}_1 \in [0, \sigma_1]$. The second and third integrals are negligible for small σ_1 (as they are both $O(\sigma_1)$ by Taylor's expansion). Combining the first and fourth integrals, I obtain

$$\int_{\sigma_1}^1 \int_{-v_1}^{1-v_1} g(r) dr dv_1 = \int_{-(1-|\sigma_1|)}^{1-|\sigma_1|} (1-|\sigma_1|-|r|) g(r) dr + O(\sigma_1)$$

and hence (by switching back to $\ell_1 = T\sigma_1$)

$$\int_{-(T-|\ell_1|)}^{T-|\ell_1|} \left(1 - \frac{|r|}{T} - \frac{|\ell_1|}{T} \right) g\left(\frac{r}{T}\right) dr + O\left(\frac{1}{T}\right).$$

The conclusion of the theorem follows as a result.

Similar to the mean calculation, the result in (A.5) implies that $(d\hat{\epsilon}_u^k)^2 = (d\epsilon_u^k)^2 + o_P(du)$, so it follows that $Var(Q) = Var(\tilde{Q})$.

A.8.3 Asymptotic normality of \tilde{Q}

The main tool for deriving asymptotic normality of \tilde{Q} is Brown's martingale central limit theorem (see, for instance, Hall and Heyde, 1980). The proof thus

boils down to three parts: (i) expressing $\tilde{Q} - E(\tilde{Q})$ as a sum of mean zero martingales, i.e. $\tilde{Q} - E(\tilde{Q}) = \sum_{i=1}^n Y_i$ where $E(Y_i|\mathcal{F}_{\tau_{i-1}}) = 0$, $n = N_T$, and τ_1, \dots, τ_n are the event times of the *pooled process* $N_t = N_t^a + N_t^b$; (ii) showing asymptotic negligibility, i.e. $s^{-4} \sum_{i=1}^n E(Y_i^4) \rightarrow 0$ where $s^2 = \text{Var}(\tilde{Q})$; and (iii) showing asymptotic determinism, i.e. $s^{-4} E(V_n^2 - s^2)^2 \rightarrow 0$, where $V_n^2 = \sum_{i=1}^n E(Y_i^2|\mathcal{F}_{\tau_{i-1}})$.

Martingale Decomposition

Recall that the statistic \tilde{Q} is defined as

$$\begin{aligned}\tilde{Q} &= \int_I w(\ell) \gamma_H^2(\ell) d\ell \\ &= \int_I w(\ell) \frac{1}{(TH)^2} \iiint_{(0,T]^4} K\left(\frac{t_1-s_1-\ell}{H}\right) K\left(\frac{t_2-s_2-\ell}{H}\right) d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_1}^b d\epsilon_{t_2}^b d\ell \\ &= \frac{1}{T^2} \iiint_{(0,T]^4} \int_I w(\ell) \frac{1}{H^2} K\left(\frac{t_1-s_1-\ell}{H}\right) K\left(\frac{t_2-s_2-\ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_1}^b d\epsilon_{t_2}^b\end{aligned}$$

I start by decomposing \tilde{Q} into four terms, corresponding to four different regions of integrations: (i) $s_1 = s_2 = s$, $t_1 = t_2 = t$; (ii) $s_1 \neq s_2$, $t_1 \neq t_2$; (iii) $s_1 \neq s_2$, $t_1 = t_2 = t$; and (iv) $s_1 = s_2 = s$, $t_1 \neq t_2$. In all cases, integrations over regions where $s_i = t_j$ for $i, j = 1, 2$ are of measure zero because of assumption (A1): the pooled point process is simple, which implies that type a and b events cannot occur at the same time almost surely. Therefore,

$$\tilde{Q} = Q_1 + Q_2 + Q_3 + Q_4 \text{ a.s.,}$$

where

$$\begin{aligned}
Q_1 &= \frac{1}{(TH)^2} \iint_{(0,T]^2} \int_I 1_{\{s \neq t\}} w(\ell) K^2\left(\frac{t-s-\ell}{H}\right) d\ell (d\epsilon_s^a)^2 (d\epsilon_t^b)^2, \\
Q_2 &= \frac{1}{(TH)^2} \iiint_{(0,T]^4} \int_I 1_{\{s_1 \neq s_2 \neq t_1 \neq t_2\}} w(\ell) K\left(\frac{t_1-s_1-\ell}{H}\right) K\left(\frac{t_2-s_2-\ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_1}^b d\epsilon_{t_2}^b, \\
Q_3 &= \frac{1}{(TH)^2} \iiint_{(0,T]^3} \int_I 1_{\{s_1 \neq s_2 \neq t\}} w(\ell) K\left(\frac{t-s_1-\ell}{H}\right) K\left(\frac{t-s_2-\ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a (d\epsilon_t^b)^2, \\
Q_4 &= \frac{1}{(TH)^2} \iiint_{(0,T]^3} \int_I 1_{\{s \neq t_1 \neq t_2\}} w(\ell) K\left(\frac{t_1-s-\ell}{H}\right) K\left(\frac{t_2-s-\ell}{H}\right) d\ell (d\epsilon_s^a)^2 d\epsilon_{t_1}^b d\epsilon_{t_2}^b.
\end{aligned}$$

I will show that (i) Q_1 contributes to the mean of \tilde{Q} ; (ii) Q_2 contributes to the variance of \tilde{Q} ; and (iii) Q_3 and Q_4 are of smaller order than Q_2 and hence asymptotically negligible.

(i) As we saw in (A.6), Q_1 is of order $O_P\left(\frac{1}{TH}\right)$ which is the largest among the four terms. I decompose Q_1 to retrieve the mean:

$$\begin{aligned}
Q_1 &= \frac{1}{(TH)^2} \iint_{(0,T]^2} \int_I w(\ell) K^2\left(\frac{t-s-\ell}{H}\right) d\ell (d\epsilon_s^a)^2 (d\epsilon_t^b)^2 \\
&= \frac{1}{(TH)^2} \iint_{(0,T]^2} \int_I w(\ell) K^2\left(\frac{t-s-\ell}{H}\right) d\ell (d\epsilon_s^a)^2 \left[(d\epsilon_t^b)^2 - \lambda_t^b dt \right] \\
&\quad + \frac{1}{(TH)^2} \iint_{(0,T]^2} \int_I w(\ell) K^2\left(\frac{t-s-\ell}{H}\right) d\ell \left[(d\epsilon_s^a)^2 - \lambda_s^a ds \right] \lambda_t^b dt \\
&\quad + \frac{1}{(TH)^2} \iint_{(0,T]^2} \int_I w(\ell) K^2\left(\frac{t-s-\ell}{H}\right) d\ell \lambda_s^a \lambda_t^b ds dt \\
&\equiv Q_{11} + Q_{12} + E(\tilde{Q}).
\end{aligned} \tag{A.8}$$

The last line is obtained by (A.6).

Lemma 22 $Q_{11} = O_P\left(\frac{1}{T^{3/2}H^{1/2}}\right)$ and $Q_{12} = O_P\left(\frac{1}{T^{3/2}H^{1/2}}\right)$ as $T \rightarrow \infty$ and $H/T \rightarrow 0$ as $H \rightarrow \infty$.

Proof. Note that Q_{11}^2 contains 5 integrals. By applying a change of variables (on

two variables inside the kernels), I deduce that $E(Q_{11}^2) = O\left(\frac{1}{T^3 H}\right)$ and hence the result. The proof for Q_{12} is similar. ■

(ii) I decompose Q_2 into $Q_2 = Q_{21} + Q_{22} + Q_{23} + Q_{24}$, where

$$\begin{aligned} Q_{21} &= \frac{1}{(TH)^2} \int_{0^+}^T \int_{0^+}^{t_2^-} \int_{0^+}^{t_2^-} \int_{0^+}^{t_2^-} \int_I 1_{\{s_1 \neq s_2 \neq t_1\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_1}^b d\epsilon_{t_2}^b \\ Q_{22} &= \frac{1}{(TH)^2} \int_{0^+}^T \int_{0^+}^{t_1^-} \int_{0^+}^{t_1^-} \int_{0^+}^{t_1^-} \int_I 1_{\{s_1 \neq s_2 \neq t_2\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_2}^b d\epsilon_{t_1}^b \\ Q_{23} &= \frac{1}{(TH)^2} \int_{0^+}^T \int_{0^+}^{s_2^-} \int_{0^+}^{s_2^-} \int_{0^+}^{s_2^-} \int_I 1_{\{t_1 \neq t_2 \neq s_1\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{t_1}^b d\epsilon_{t_2}^b d\epsilon_{s_1}^a d\epsilon_{s_2}^a \\ Q_{24} &= \frac{1}{(TH)^2} \int_{0^+}^T \int_{0^+}^{s_1^-} \int_{0^+}^{s_1^-} \int_{0^+}^{s_1^-} \int_I 1_{\{t_1 \neq t_2 \neq s_2\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{t_1}^b d\epsilon_{t_2}^b d\epsilon_{s_2}^a d\epsilon_{s_1}^a \end{aligned}$$

Lemma 23 $Q_2 = O_P\left(\frac{1}{TH}\right) + O_P\left(\frac{1}{TH^{1/2}}\right)$ as $T \rightarrow \infty$ and $H/T \rightarrow 0$ as $H \rightarrow \infty$.

Proof. Indeed, the asymptotic variance of \tilde{Q} in (A.7) comes from Q_2 . ■

(iii) It turns out that Q_3 and Q_4 are asymptotically negligible compared to Q_2 .

Lemma 24 $Q_3 = O_P\left(\frac{1}{T^{3/2}H^{1/2}}\right)$ and $Q_4 = O_P\left(\frac{1}{T^{3/2}H^{1/2}}\right)$ as $T \rightarrow \infty$ and $H/T \rightarrow 0$ as $H \rightarrow \infty$.

Proof. Note that Q_3^2 contains 5 integrals. By applying a change of variables (on three variables inside the kernels) and combining $w(\ell_1)$ and $w(\ell_2)$ into $w^2(\ell)$ in the process, we deduce that $E(Q_3^2) = O\left(\frac{1}{T^3 H}\right)$ and hence the result. The proof for Q_4 is similar. ■

As a result,

$$\tilde{Q} - E(\tilde{Q}) = Q_2 + O_P\left(\frac{1}{T^{3/2}H^{1/2}}\right).$$

Now, I want to show that Q_2 , the leading term of the demeaned statistic, can be expressed into the sum of a martingale difference sequence (m.d.s.).

Lemma 25 *Let $n = N(T)$ be the total event counts of the pooled process $N = N^a + N^b$. Then, as $T \rightarrow \infty$ and $H/T \rightarrow 0$ as $H \rightarrow \infty$.*

$$\tilde{Q} - E(\tilde{Q}) = \sum_{i=1}^n Y_i + O_P\left(\frac{1}{T^{3/2}H^{1/2}}\right)$$

where $Y_i = \sum_{j=1}^4 Y_{ji}$ and $E(Y_{ji}|\mathcal{F}_{\tau_{i-1}}^{ab}) = 0$ for all $i = 1, \dots, n$ and for $j = 1, 2, 3, 4$ (i.e. $\{Y_{ji}\}_{i=1}^n$ are m.d.s. for $j = 1, 2, 3, 4$).

Proof. The result follows by defining

$$\begin{aligned} Y_{1i} &= \frac{1}{(TH)^2} \int_{\tau_{i-1}^+}^{\tau_i} \int_0^{\tau_2^-} \int_0^{\tau_2^-} \int_I 1_{\{s_1 \neq s_2 \neq t_1\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_1}^b d\epsilon_{t_2}^b, \\ Y_{2i} &= \frac{1}{(TH)^2} \int_{\tau_{i-1}^+}^{\tau_i} \int_0^{\tau_1^-} \int_0^{\tau_1^-} \int_I 1_{\{s_1 \neq s_2 \neq t_2\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_2}^b d\epsilon_{t_1}^b, \\ Y_{3i} &= \frac{1}{(TH)^2} \int_{\tau_{i-1}^+}^{\tau_i} \int_0^{\tau_2^-} \int_0^{\tau_2^-} \int_I 1_{\{s_1 \neq t_1 \neq t_2\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{t_2}^b d\epsilon_{t_1}^b d\epsilon_{s_2}^a, \\ Y_{4i} &= \frac{1}{(TH)^2} \int_{\tau_{i-1}^+}^{\tau_i} \int_0^{\tau_1^-} \int_0^{\tau_1^-} \int_I 1_{\{s_2 \neq t_1 \neq t_2\}} w(\ell) K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{s_2}^a d\epsilon_{t_2}^b d\epsilon_{t_1}^b d\epsilon_{s_1}^a. \end{aligned}$$

and noting that $E(Y_{ji}|\mathcal{F}_{\tau_{i-1}}^{ab}) = 0$ for all $i = 1, \dots, n$. ■

Asymptotic Negligibility

Next, I want to show that the summation $\sum_{i=1}^n Y_i^4$ is asymptotically negligible compared to $[Var(\tilde{Q})]^2$.

Lemma 26 $s^{-4} \sum_{i=1}^n E(Y_i^4) \rightarrow 0$ as $T \rightarrow \infty$ and $H/T \rightarrow 0$ as $H \rightarrow \infty$, where $s^2 = Var(\tilde{Q})$.

Proof. Consider

$$Y_{li}^4 = \frac{1}{T^8 H^8} \iiint_{(\tau_{i-1}, \tau_i]^4} \int_{(0, t_2)^{12}} \cdots \int_{I^4} w(\ell_1) \dots w(\ell_4) K\left(\frac{t_{111} - s_{111} - \ell_1}{H}\right) \dots$$

$$K\left(\frac{t_{222} - s_{222} - \ell_4}{H}\right) d\ell_1 \dots d\ell_4 d\epsilon_{s_{111}}^a \dots d\epsilon_{s_{222}}^a d\epsilon_{t_{111}}^b \dots d\epsilon_{t_{222}}^b.$$

A key observation is that $t_{211} = t_{212} = t_{221} = t_{222} \equiv t_2$ because there is at most one event of type b in the interval $(\tau_{i-1}, \tau_i]$ (one event if τ_i is a type b event time, zero events if τ_i is a type a event time). This reduces the four outermost integrations to just one over $t_2 \in (\tau_{i-1}, \tau_i]$. Let us focus on extracting the dominating terms. Then, to maximize the order of magnitude of $E(Y_{li}^4)$, the next 12 integrations can be reduced to six integrations after grouping $d\epsilon_{ijl}^a$ and $d\epsilon_{1jl}^b$ into six pairs (if they were not paired, then the corresponding contribution to $E(Y_{li}^4)$ would be zero by iterated expectations). Together with the four innermost integrations, there are 11 integrations for Y_{11i}^4 , with the outermost integration running over $(\tau_{i-1}, \tau_i]$. Therefore, there are 11 integrations in $\sum_{i=1}^n E(Y_{li}^4)$ and its outermost integration with respect to t_2 runs over $(0, T]$. As six new variables are sufficient to represent all 12 arguments in the 12 kernels, a change of variables yields a factor of $TH^6 \kappa_4^2$.³ As a result, $\sum_{i=1}^n E(Y_{li}^4) = O\left(\frac{1}{T^7 H^2}\right)$, and since $s^2 = O\left(\frac{1}{T^2 H}\right)$ from (A.7), we have $s^{-4} \sum_{i=1}^n E(Y_{li}^4) = O\left(\frac{1}{T^3}\right)$. The same argument applies to Y_{ji} for $j = 2, 3, 4$. By Minkowski's inequality $s^{-4} \sum_{i=1}^n E(Y_i^4) = O\left(\frac{1}{T^3}\right)$. ■

Asymptotic Determinism

Lastly, I want to show that the variance of $V_n^2 = \sum_{i=1}^n E(Y_i^2 | \mathcal{F}_{\tau_{i-1}})$ is of a smaller order than s^4 .

³11 integrations - 6 d.f. - 4 $w()$ = 1 free integration with respect to t .

Lemma 27 $s^{-4}E(V_n^2 - s^2)^2 \rightarrow 0$ as $T \rightarrow \infty$ and $H/T \rightarrow 0$ as $H \rightarrow \infty$.

Proof. To prove that $s^{-4}E(V_n^2 - s^2)^2 \rightarrow 0$, it suffices to show that

$$E(V_n^2 - s^2)^2 = o\left(\frac{1}{T^4 H^2}\right). \quad (\text{A.9})$$

(i) Recall from lemma 25 that the i^{th} term of the martingale difference sequence in the demeaned statistic $\tilde{Q} - E(\tilde{Q})$ represents the innovation in the time interval $(\tau_{i-1}, \tau_i]$ and is given by $Y_i = Y_{1i} + Y_{2i} + Y_{3i} + Y_{4i}$, for $i = 1, 2, \dots, n = N(T)$.

Now, note that

$$Y_i^2 = Y_{1i}^2 + Y_{2i}^2 + Y_{3i}^2 + Y_{4i}^2 + 2Y_{1i}Y_{2i} + 2Y_{3i}Y_{4i} \quad (\text{A.10})$$

almost surely. The terms $Y_{1i}Y_{3i}$, $Y_{1i}Y_{4i}$, $Y_{2i}Y_{3i}$ and $Y_{2i}Y_{4i}$ are almost surely zero because of assumption (A1): the pooled process $N = N^a + N^b$ is simple, which implies that type a and b events will not occur at the same time τ_i almost surely.

(ii) Define

$$S_1 = S_2 \equiv \frac{1}{T^4 H^4} \int_{0^+}^T \iiint_{(0, t_2)^3} \iint_{I^2} w(\ell_1) w(\ell_2) K\left(\frac{t_1 - s_{11} - \ell_1}{H}\right) K\left(\frac{t_1 - s_{12} - \ell_2}{H}\right) \\ K\left(\frac{t_2 - s_{21} - \ell_1}{H}\right) K\left(\frac{t_2 - s_{22} - \ell_2}{H}\right) 1_{\mathcal{R}_1 \cup \mathcal{R}_4 \cup \mathcal{R}_7} d\ell_1 d\ell_2 \lambda_{s_1}^a \lambda_{s_2}^a \lambda_{t_1}^b \lambda_{t_2}^b ds_1 ds_2 dt_1 dt_2,$$

$$S_3 = S_4 \equiv \frac{1}{T^4 H^4} \int_{0^+}^T \iiint_{(0, s_2)^3} \iint_{I^2} w(\ell_1) w(\ell_2) K\left(\frac{t_{11} - s_1 - \ell_1}{H}\right) K\left(\frac{t_{12} - s_1 - \ell_2}{H}\right) \\ K\left(\frac{t_{21} - s_2 - \ell_1}{H}\right) K\left(\frac{t_{22} - s_2 - \ell_2}{H}\right) 1_{\mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3} d\ell_1 d\ell_2 \lambda_{s_1}^a \lambda_{s_2}^a \lambda_{t_1}^b \lambda_{t_2}^b ds_1 dt_1 dt_2 ds_2,$$

$$S_{12} \equiv \frac{1}{T^4 H^4} \int_{0^+}^T \iiint_{(0, t_2)^3} \iint_{I^2} w(\ell_1) w(\ell_2) K\left(\frac{t_1 - s_{11} - \ell_1}{H}\right) K\left(\frac{t_1 - s_{22} - \ell_2}{H}\right) \\ K\left(\frac{t_2 - s_{21} - \ell_1}{H}\right) K\left(\frac{t_2 - s_{12} - \ell_2}{H}\right) 1_{\mathcal{R}_3 \cup \mathcal{R}_6 \cup \mathcal{R}_9} d\ell_1 d\ell_2 \lambda_{s_1}^a \lambda_{s_2}^a \lambda_{t_1}^b \lambda_{t_2}^b ds_1 ds_2 dt_1 dt_2,$$

and

$$S_{34} \equiv \frac{1}{T^4 H^4} \int_{0^+}^T \iiint_{(0, s_2)^3} \iint_{I^2} w(\ell_1) w(\ell_2) K\left(\frac{t_{11}-s_1-\ell_1}{H}\right) K\left(\frac{t_{22}-s_1-\ell_2}{H}\right) \\ K\left(\frac{t_{21}-s_2-\ell_1}{H}\right) K\left(\frac{t_{12}-s_2-\ell_2}{H}\right) 1_{\mathcal{R}_7 \cup \mathcal{R}_8 \cup \mathcal{R}_9} d\ell_1 d\ell_2 \lambda_{s_1}^a \lambda_{s_2}^a \lambda_{t_1}^b \lambda_{t_2}^b dt_1 dt_2 ds_1 ds_2,$$

where \mathcal{R}_i were defined in section A.8.2. It is easy to verify from the definitions that

$$s^2 = S_1 + S_2 + S_3 + S_4 + 2S_{12} + 2S_{34} + o_p\left(\frac{1}{T^2 H}\right). \quad (\text{A.11})$$

(iii) Define for $k = 1, 2, 3, 4$

$$V_{nk}^2 \equiv \sum_{i=1}^n E(Y_{ki}^2 | \mathcal{F}_{\tau_{i-1}}),$$

and for $(k, j) = (1, 2)$ and $(3, 4)$

$$V_{nkj} \equiv \sum_{i=1}^n E(Y_{ki} Y_{ji} | \mathcal{F}_{\tau_{i-1}}).$$

It follows from (A.10) that

$$V_n^2 = V_{1n}^2 + V_{2n}^2 + V_{3n}^2 + V_{4n}^2 + 2V_{12n} + 2V_{34n}. \quad (\text{A.12})$$

(iv) I claim (see the proof below) that, for $k = 1, 2, 3, 4$,

$$V_{nk}^2 - S_k = o_p\left(\frac{1}{T^2 H}\right) \quad (\text{A.13})$$

and, for $(k, j) = (1, 2)$ and $(3, 4)$,

$$V_{nkj}^2 - S_{kj} = o_p\left(\frac{1}{T^2 H}\right). \quad (\text{A.14})$$

(v) It follows from (A.10)-(A.14) that (A.9) holds.

It remains to show the claims in (iv). Since the asymptotic orders of the six differences in (A.13) and (A.14) can be derived by similar techniques, let us focus on proving the first one.

To this end, I first compute $E(Y_{1i}^2 | \mathcal{F}_{\tau_{i-1}})$. Now,

$$Y_{1i}^2 = \frac{1}{T^4 H^4} \iint_{(\tau_{i-1}, \tau_i]^2} \int \cdots \int_{(0, t_2)^6} \iint_{I^2} w(\ell_1) w(\ell_2) K\left(\frac{t_{11} - s_{11} - \ell_1}{H}\right) K\left(\frac{t_{12} - s_{12} - \ell_2}{H}\right) \\ K\left(\frac{t_2 - s_{21} - \ell_1}{H}\right) K\left(\frac{t_2 - s_{22} - \ell_2}{H}\right) d\ell_1 d\ell_2 d\epsilon_{s_{11}}^a d\epsilon_{s_{12}}^a d\epsilon_{s_{21}}^a d\epsilon_{s_{22}}^a d\epsilon_{t_{11}}^b d\epsilon_{t_{12}}^b d\epsilon_{t_{21}}^b d\epsilon_{t_{22}}^b.$$

Observe that there is at most one event of type b in the interval $(\tau_{i-1}, \tau_i]$ (one event if τ_i is a type b event time, zero events if τ_i is a type a event time). This entails that $t_{21} = t_{22} \equiv t_2$ and thus saves one integration. I can then rewrite

$$Y_{1i}^2 = \frac{1}{T^4 H^4} \iint_{(\tau_{i-1}, \tau_i]} \int \cdots \int_{(0, t_2)^6} \iint_{I^2} w(\ell_1) w(\ell_2) K\left(\frac{t_{11} - s_{11} - \ell_1}{H}\right) K\left(\frac{t_{12} - s_{12} - \ell_2}{H}\right) \\ K\left(\frac{t_2 - s_{21} - \ell_1}{H}\right) K\left(\frac{t_2 - s_{22} - \ell_2}{H}\right) d\ell_1 d\ell_2 d\epsilon_{s_{11}}^a d\epsilon_{s_{12}}^a d\epsilon_{s_{21}}^a d\epsilon_{s_{22}}^a d\epsilon_{t_{11}}^b d\epsilon_{t_{12}}^b (d\epsilon_{t_2}^b)^2 \\ \equiv \int_{(\tau_{i-1}, \tau_i]} H_{11}(t_2^-) (d\epsilon_{t_2}^b)^2,$$

where I define $H_{11}(u)$ by

$$H_{11}(u^-) \equiv \int \cdots \int_{(0, u)^6} \iint_{I^2} w(\ell_1) w(\ell_2) K\left(\frac{t_{11} - s_{11} - \ell_1}{H}\right) K\left(\frac{t_{12} - s_{12} - \ell_2}{H}\right) K\left(\frac{u - s_{21} - \ell_1}{H}\right) \\ K\left(\frac{u - s_{22} - \ell_2}{H}\right) d\ell_1 d\ell_2 d\epsilon_{s_{11}}^a d\epsilon_{s_{12}}^a d\epsilon_{s_{21}}^a d\epsilon_{s_{22}}^a d\epsilon_{t_{11}}^b d\epsilon_{t_{12}}^b.$$

Note that $H_{11}(u^-)$ is \mathcal{F} -predictable. Now, by iterated expectations, lemma 1, and the fact that $\{u \in (\tau_{i-1}, \tau_i]\} \in \mathcal{F}_{u^-}$, I have

$$E(Y_{1i}^2 | \mathcal{F}_{\tau_{i-1}}) = E \left\{ \int_{\tau_{i-1}^+}^{\tau_i} H_{11}(u^-) (d\epsilon_u^b)^2 \middle| \mathcal{F}_{\tau_{i-1}} \right\} \\ = E \left\{ \int_{\tau_{i-1}^+}^{\tau_i} H_{11}(u^-) \lambda_u^b du \middle| \mathcal{F}_{\tau_{i-1}} \right\} \\ = E \left\{ \int_{\tau_{i-1}^+}^{\tau_i} H_{11}(u^-) \frac{\lambda_u^b}{\lambda_u^a + \lambda_u^b} dN_u \middle| \mathcal{F}_{\tau_{i-1}} \right\} \\ = H_{11}(\tau_{i-1}) E \left\{ \frac{\lambda_{\tau_i}^b}{\lambda_{\tau_i}^a + \lambda_{\tau_i}^b} \middle| \mathcal{F}_{\tau_{i-1}} \right\}.$$

Note that I used the property $H_{11}(\tau_i^-) = H_{11}(\tau_{i-1})$ in the last line. Summing over i gives

$$\begin{aligned}
V_{n1}^2 &\equiv \sum_{i=1}^n E(Y_{1i}^2 | \mathcal{F}_{\tau_{i-1}}) \\
&= \sum_{i=1}^n H_{11}(\tau_{i-1}) E \left\{ \frac{\lambda_{\tau_i}^b}{\lambda_{\tau_i}^a + \lambda_{\tau_i}^b} \middle| \mathcal{F}_{\tau_{i-1}} \right\} \\
&= \int_0^T H_{11}(u^-) E \left\{ \frac{\lambda_u^b}{\lambda_u^a + \lambda_u^b} \middle| \mathcal{F}_{u^-} \right\} dN_u \\
&= \int_0^T H_{11}(u^-) \frac{\lambda_u^b}{\lambda_u^a + \lambda_u^b} dN_u.
\end{aligned}$$

The third equality made use of the property that for $u \in (\tau_{i-1}, \tau_i]$, $N_{u^-} = N_{\tau_{i-1}}$ and hence $\mathcal{F}_{u^-} = \sigma\{(\tau_i, y_i) : 0 \leq i \leq t_{N_{u^-}}\} = \mathcal{F}_{\tau_{i-1}}$, and the last line follows from \mathcal{F}_t -predictability of conditional intensities λ_t^a and λ_t^b .

Let $\pi_u^b = \frac{\lambda_u^b}{\lambda_u^a + \lambda_u^b}$. Apart from the terms with $t_{11} \neq t_{12}$ and/or $s_{ij} \neq s_{kl}$ for $(i, j) \neq (k, l)$ which can be shown to be $O_p\left(\frac{1}{T^6 H^2}\right) = o_p\left(\frac{1}{T^4 H^2}\right)$, the integral $V_{n1}^2 = \int_0^T H_{11}(t_2^-) \pi_{t_2}^b dN_{t_2}$ can be decomposed by the same demeaning technique as we used for decomposing Q_1 in (A.8). The decomposition is represented by differentials for simplicity:

$$\begin{aligned}
&d\ell_1 d\ell_2 \left(d\epsilon_{s_1}^a \right)^2 \left(d\epsilon_{s_2}^a \right)^2 \left(d\epsilon_{t_1}^b \right)^2 \pi_{t_2}^b dN_{t_2} \\
&= d\ell_1 d\ell_2 \left(d\epsilon_{s_1}^a \right)^2 \left(d\epsilon_{s_2}^a \right)^2 \left(d\epsilon_{t_1}^b \right)^2 \pi_{t_2}^b \left[dN_{t_2} - (\lambda_{t_2}^a + \lambda_{t_2}^b) dt_2 \right] \\
&+ d\ell_1 d\ell_2 \left(d\epsilon_{s_1}^a \right)^2 \left(d\epsilon_{s_2}^a \right)^2 \left[\left(d\epsilon_{t_1}^b \right)^2 - \lambda_{t_1}^b dt_1 \right] \lambda_{t_2}^b dt_2 \\
&+ d\ell_1 d\ell_2 \left(d\epsilon_{s_1}^a \right)^2 \left[\left(d\epsilon_{s_2}^a \right)^2 - \lambda_{s_2}^a ds_2 \right] \lambda_{t_1}^b \lambda_{t_2}^b dt_1 dt_2 \\
&+ d\ell_1 d\ell_2 \left[\left(d\epsilon_{s_1}^a \right)^2 - \lambda_{s_1}^a ds_1 \right] \lambda_{s_2}^a \lambda_{t_1}^b \lambda_{t_2}^b ds_2 dt_1 dt_2 \\
&+ d\ell_1 d\ell_2 \lambda_{s_1}^a \lambda_{s_2}^a \lambda_{t_1}^b \lambda_{t_2}^b ds_1 ds_2 dt_1 dt_2.
\end{aligned}$$

The first four integrals above are dominated by the first term, which can be shown to be of size $O_p\left(\frac{1}{(T^5 H^2)^{1/2}}\right) = o_p\left(\frac{1}{(T^4 H^2)^{1/2}}\right)$. The last integral is S_1 which

contributes to $s^2 = \text{Var}(\tilde{Q})$ and was proven to be $O_p\left(\frac{1}{(T^4 H^2)^{1/2}}\right)$ in Theorem 8. Hence, $V_{n1}^2 - S_1 = o_p\left(\frac{1}{T^2 H}\right)$. ■

A.9 Proof of Theorem 10

First, recall that

$$\tilde{Q} = \frac{1}{T^2} \iiint_{(0,T]^4} \int_I w(\ell) \frac{1}{H^2} K\left(\frac{t_1 - s_1 - \ell}{H}\right) K\left(\frac{t_2 - s_2 - \ell}{H}\right) d\ell d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_1}^b d\epsilon_{t_2}^b.$$

From the property of the joint cumulant of the innovations, all of which have mean zero, I can express

$$\begin{aligned} E[d\epsilon_{s_1}^a d\epsilon_{s_2}^a d\epsilon_{t_1}^b d\epsilon_{t_2}^b] &= E[d\epsilon_{s_1}^a d\epsilon_{s_2}^a] E[d\epsilon_{t_1}^b d\epsilon_{t_2}^b] + E[d\epsilon_{s_1}^a d\epsilon_{t_1}^b] E[d\epsilon_{s_2}^a d\epsilon_{t_2}^b] \\ &\quad + E[d\epsilon_{s_1}^a d\epsilon_{t_2}^b] E[d\epsilon_{s_2}^a d\epsilon_{t_1}^b] + c_{22}(s_2 - s_1, t_1 - s_1, t_2 - s_1) \\ &= 0 + \gamma(t_1 - s_1)\gamma(t_2 - s_2) + \gamma(t_2 - s_1)\gamma(t_1 - s_2) \\ &\quad + c_{22}(s_2 - s_1, t_1 - s_1, t_2 - s_1) \\ &= a_T^2 \lambda^a \lambda^b [\rho(t_1 - s_1)\rho(t_2 - s_2) + \rho(t_2 - s_1)\rho(t_1 - s_2)] \\ &\quad + o(a_T^2), \end{aligned}$$

where the last line utilizes assumption (A8).

Since $H = o(B)$, the asymptotic bias of \tilde{Q} becomes

$$\text{bias}(\tilde{Q}) = \frac{a_T^2 \lambda^a \lambda^b}{TH} \kappa_2 \int_I w_B(\ell) \left(1 - \frac{|\ell|}{T}\right) \check{\rho}^2(\ell) d\ell + o\left(\frac{a_T^2}{TH}\right).$$

The asymptotic variance of \tilde{Q} under \mathbf{H}_{a_T} is the same as that under \mathbf{H}_0 and was given in Theorem 8. If I set $a_T^* = H^{1/4}$, then the normalized statistic J converges in distribution to $N(\mu(K, w_B), 1)$.

A.10 Proof of Theorem 9

I will only prove the case with no autocorrelations, i.e. $c^{kk}(\ell) \equiv 0$ for $k = a, b$, as the error of estimating auto-covariances by their estimators can be made negligible by similar techniques as in the case for conditional intensities.

First, assuming the setup in section 2.4.4, the conditional intensity λ_t^k can be approximated by $\hat{\lambda}_t^k - \lambda_t^k = O_P(M^{-1/2})$ by Theorem 6.

Next, by Theorem 6, it follows that $\hat{\lambda}_t^k - \lambda_t^k = O_P(M^{-1/2})$ for $k = a, b$. By lemma 28 (see below), it is true that $T(Q - \tilde{Q}) = O_P(M^{-1/2})$. By the assumption $H = o(M)$, I thus obtain $T(Q - \tilde{Q}) = o_P(H^{-1/2})$, and hence, with $\text{Var}(TQ) = O(H^{-1/2})$,

$$T(Q - \tilde{Q}) / \sqrt{\text{Var}(TQ)} = o_P(1). \quad (\text{A.15})$$

Besides, note that the approximation error of the unconditional intensity $\hat{\lambda}^k - \lambda^k$ is diminishing at the parametric rate of $O_P(T^{-1/2}) = o_P(1)$ as $T \rightarrow \infty$. Also, note that $\widehat{E(TQ)}$ is a function of unconditional intensities, so (A.6) implies that $\widehat{E(TQ)} - E(TQ) = o(H^{-1})$, or

$$\left[\widehat{E(TQ)} - E(TQ) \right] / \sqrt{\text{Var}(TQ)} = o(H^{-1/2}) = o(1). \quad (\text{A.16})$$

Furthermore, the estimated variance $\widehat{\text{Var}(TQ)}$ is a function of unconditional intensities too, so (A.7) implies that $\widehat{\text{Var}(TQ)} - \text{Var}(TQ) = o(H^{-1})$, or

$$\text{Var}(TQ) / \widehat{\text{Var}(TQ)} = 1 + o(1). \quad (\text{A.17})$$

Lastly, the result follows from the decomposition below with an application

of Slutsky's theorem, meanwhile making use of (A.15), (A.16) and (A.17):

$$\begin{aligned}
\hat{J} &= \frac{TQ - E(\widehat{TQ})}{\sqrt{\widehat{\text{Var}}(TQ)}} \\
&= \sqrt{\frac{\text{Var}(TQ)}{\widehat{\text{Var}}(TQ)}} \left[\frac{T\tilde{Q} - E(TQ)}{\sqrt{\text{Var}(TQ)}} + \frac{T(Q - \tilde{Q})}{\sqrt{\text{Var}(TQ)}} + \frac{E(\widehat{TQ}) - E(TQ)}{\sqrt{\text{Var}(TQ)}} \right] \\
&= J + o_P(1).
\end{aligned}$$

Lemma 28 $T(Q - \tilde{Q}) = O_P(M^{-1/2})$.

Proof. Recall the statistic $Q = \int_{I/T} w_b(\sigma) \widehat{\gamma}_h^2(\sigma) d\sigma$ and its hypothetical counterpart $\tilde{Q} = \int_{I/T} w_b(\sigma) \tilde{\gamma}_h^2(\sigma) d\sigma$. ■

To evaluate the asymptotic order of $Q - \tilde{Q}$, I apply a change of variable $s = Tu$ and $t = Tv$ as described in section 2.4.6. From Theorem 6, we know that $\hat{\lambda}_v^k - \tilde{\lambda}_v^k = \hat{\lambda}_{T_v}^k - \lambda_{T_v}^k = O_P(M^{-1/2})$ for $k = a, b$ (u and v fixed). It follows that $\hat{\lambda}_u^a \hat{\lambda}_v^b - \tilde{\lambda}_u^a \tilde{\lambda}_v^b = (\hat{\lambda}_u^a - \tilde{\lambda}_u^a) \hat{\lambda}_v^b + (\hat{\lambda}_v^b - \tilde{\lambda}_v^b) \tilde{\lambda}_u^a = O_P(M^{-1/2})$, and hence

$$\begin{aligned}
d\hat{\varepsilon}_u^a d\hat{\varepsilon}_v^b - d\tilde{\varepsilon}_u^a d\tilde{\varepsilon}_v^b &= (d\tilde{N}_u^a - \hat{\lambda}_u^a du)(d\tilde{N}_v^b - \hat{\lambda}_v^b dv) - (d\tilde{N}_u^a - \tilde{\lambda}_u^a du)(d\tilde{N}_v^b - \tilde{\lambda}_v^b dv) \\
&= -d\tilde{N}_u^a (\hat{\lambda}_v^b - \tilde{\lambda}_v^b) dv - d\tilde{N}_v^b (\hat{\lambda}_u^a - \tilde{\lambda}_u^a) du + (\hat{\lambda}_u^a \hat{\lambda}_v^b - \tilde{\lambda}_u^a \tilde{\lambda}_v^b) dudv \\
&= O_P(M^{-1/2}).
\end{aligned}$$

As a result, $\widehat{\gamma}_h(\sigma) - \tilde{\gamma}_h(\sigma) = \int_0^1 \int_0^1 K_h(v - u - \sigma) [d\widehat{\varepsilon}_u^a d\widehat{\varepsilon}_v^b - d\tilde{\varepsilon}_u^a d\tilde{\varepsilon}_v^b] = O_P(M^{-1/2})$.

Since $\widehat{\gamma}_h(\sigma) + \tilde{\gamma}_h(\sigma) = O_P(1)$, I deduce that

$$\begin{aligned}
\widehat{\gamma}_h^2(\sigma) - \tilde{\gamma}_h^2(\sigma) &= (\widehat{\gamma}_h(\sigma) + \tilde{\gamma}_h(\sigma)) (\widehat{\gamma}_h(\sigma) - \tilde{\gamma}_h(\sigma)) \\
&= O_P(M^{-1/2}),
\end{aligned}$$

and thus conclude that $Q - \tilde{Q} = \int_{I/T} w_b(\sigma) [\widehat{\gamma}_h^2(\sigma) - \tilde{\gamma}_h^2(\sigma)] d\sigma = O_P(T^{-1} M^{-1/2})$,

which implies that $T(Q - \tilde{Q}) = O_P(M^{-1/2})$.

A.11 Proof of Corollary 11

It suffices to show that the mean and variance are as given in the corollary. Denote the delta function by $\delta(\cdot)$. Since $B = o(H)$ as $H \rightarrow \infty$, the following approximation is valid:

$$\begin{aligned} & \frac{1}{B} w\left(\frac{\ell}{B}\right) \frac{1}{H^2} K\left(\frac{u-\ell}{H}\right) K\left(\frac{v-\ell}{H}\right) \\ &= \delta_\ell(0) \frac{1}{H^2} K\left(\frac{u-\ell}{H}\right) K\left(\frac{v-\ell}{H}\right) + o(1). \end{aligned}$$

Therefore,

$$\begin{aligned} Q &= \int_I w_B(\ell) \hat{\gamma}_H^2(\ell) d\ell \\ &= \frac{1}{T^2} \iiint_{(0,T]^4} \int_0^T \frac{1}{B} w\left(\frac{\ell}{B}\right) \frac{1}{H^2} K\left(\frac{t_1-s_1-\ell}{H}\right) K\left(\frac{t_2-s_2-\ell}{H}\right) d\ell d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b \\ &= \frac{1}{T^2} \iint_{(0,T]^2} \int_{t_2-T}^{t_2} \int_{t_1-T}^{t_1} \int_0^T \frac{1}{B} w\left(\frac{\ell}{B}\right) \frac{1}{H^2} K\left(\frac{u-\ell}{H}\right) K\left(\frac{v-\ell}{H}\right) d\ell d\hat{\epsilon}_{t_1-u}^a d\hat{\epsilon}_{t_2-v}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b \\ &= \frac{1}{T^2} \iint_{(0,T]^2} \int_{t_2-T}^{t_2} \int_{t_1-T}^{t_1} \left[\frac{1}{H^2} K\left(\frac{u}{H}\right) K\left(\frac{v}{H}\right) + o(1) \right] d\hat{\epsilon}_{t_1-u}^a d\hat{\epsilon}_{t_2-v}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b. \end{aligned}$$

Under the null hypothesis (2.11), I compute the mean (up to the leading term) as follows:

$$\begin{aligned} E(Q) &= \frac{1}{T^2} \iint_{(0,T]^2} \int_{t_2-T}^{t_2} \int_{t_1-T}^{t_1} \frac{1}{H^2} K\left(\frac{u}{H}\right) K\left(\frac{v}{H}\right) E\left(d\hat{\epsilon}_{t_1-u}^a d\hat{\epsilon}_{t_2-v}^a\right) E\left(d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b\right) \\ &= \frac{\lambda^a \lambda^b}{T^2} \int_0^T \int_{t-T}^t \frac{1}{H^2} K^2\left(\frac{u}{H}\right) du dt. \end{aligned}$$

By Fubini's theorem, the last line becomes

$$\begin{aligned} E(Q) &= \frac{\lambda^a \lambda^b}{T} \int_{-T}^T \frac{1}{H^2} K^2\left(\frac{u}{H}\right) \left(1 - \frac{|u|}{T}\right) du \\ &= \frac{\lambda^a \lambda^b}{TH} \int_{-T/h}^{T/h} K^2(v) \left(1 - \frac{|v|H}{T}\right) dv \\ &= \frac{\lambda^a \lambda^b}{TH} [k_2 + o(1)]. \end{aligned}$$

so that $E(Q^{\mathcal{G}}) = C^{\mathcal{G}} + o(1)$. By similar techniques as I obtained (A.7), I compute the second moment (up to the leading term) as follows.

$$E(Q^2) = \frac{1}{(TH)^4} E \left[\prod_{i,j=1,2} \int_0^T \int_{t_{ij}-T}^{t_{ij}} K\left(\frac{u_{ij}}{H}\right) d\tilde{\epsilon}_{t_{ij}-u_{ij}}^a d\tilde{\epsilon}_{t_{ij}}^b \right]$$

The leading order terms of $E(Q^2)$ are obtained when:

- (1) $t_{11} = t_{12}, t_{21} = t_{22}, t_{11} - u_{11} = t_{21} - u_{21}, t_{12} - u_{12} = t_{22} - u_{22};$
- (2) $t_{11} = t_{12}, t_{21} = t_{22}, t_{11} - u_{11} = t_{22} - u_{22}, t_{12} - u_{12} = t_{21} - u_{21};$
- (3) $t_{11} = t_{21}, t_{12} = t_{22}, t_{11} - u_{11} = t_{12} - u_{12}, t_{21} - u_{21} = t_{22} - u_{22};$
- (4) $t_{11} = t_{21}, t_{12} = t_{22}, t_{11} - u_{11} = t_{22} - u_{22}, t_{12} - u_{12} = t_{21} - u_{21};$
- (5) $t_{11} = t_{22}, t_{12} = t_{21}, t_{11} - u_{11} = t_{12} - u_{12}, t_{21} - u_{21} = t_{22} - u_{22};$
- (6) $t_{11} = t_{22}, t_{12} = t_{21}, t_{11} - u_{11} = t_{21} - u_{21}, t_{12} - u_{12} = t_{22} - u_{22}.$

Their contributions add up to

$$\frac{6(\lambda^a \lambda^b)^2}{(TH)^4} \iint_{(0,T]^2} \int_{t_2-T}^{t_2} \int_{t_1-T}^{t_1} \int_A K\left(\frac{u_1}{H}\right) K\left(\frac{u_2}{H}\right) K\left(\frac{u_1+v}{H}\right) K\left(\frac{u_2+v}{H}\right) dv du_1 du_2 dt_1 dt_2$$

where $A = \cap_{i=1}^2 [t_i - T - u_i, t_i - u_i]$. After a change of variables, the last line reduces to

$$\frac{6(\lambda^a \lambda^b)^2}{T^2 H} [\kappa_4 + o(1)],$$

which dominates $[E(Q)]^2$. As a result, $\text{Var}(Q^{\mathcal{G}}) = 2D^{\mathcal{G}} + o(1)$.

A.12 Proof of Corollary 12

It suffices to show that the mean and variance are as given in the corollary. Denote the Dirac delta function at ℓ by $\delta_\ell(\cdot)$. Since $H = o(B)$ as $B \rightarrow \infty$, the following approximation is valid:

$$\frac{1}{B}w\left(\frac{\ell}{B}\right)\frac{1}{H^2}K\left(\frac{u-\ell}{H}\right)K\left(\frac{v-\ell}{H}\right) = \delta_\ell(u)\delta_\ell(v)\frac{1}{B}w\left(\frac{\ell}{B}\right) + o(1).$$

Therefore,

$$\begin{aligned} Q &= \int_I w_B(\ell) \hat{\gamma}_H^2(\ell) d\ell \\ &= \frac{1}{T^2} \iiint_{(0,T]^4} \int_0^T \frac{1}{B}w\left(\frac{\ell}{B}\right)\frac{1}{H^2}K\left(\frac{t_1-s_1-\ell}{H}\right)K\left(\frac{t_2-s_2-\ell}{H}\right) d\ell d\hat{\epsilon}_{s_1}^a d\hat{\epsilon}_{s_2}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b \\ &= \frac{1}{T^2} \iint_{(0,T]^2} \int_{t_2-T}^{t_2} \int_{t_1-T}^{t_1} \int_0^T \frac{1}{B}w\left(\frac{\ell}{B}\right)\frac{1}{H^2}K\left(\frac{u-\ell}{H}\right)K\left(\frac{v-\ell}{H}\right) d\ell d\hat{\epsilon}_{t_1-u}^a d\hat{\epsilon}_{t_2-v}^a d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b \\ &= \frac{1}{T^2} \iint_{(0,T]^2} \int_0^T 1_{\{\ell \in \cap_{i=1}^2 [t_i-T, t_i]\}} \left[\frac{1}{B}w\left(\frac{\ell}{B}\right) + o(1) \right] d\hat{\epsilon}_{t_1-\ell}^a d\hat{\epsilon}_{t_2-\ell}^a d\ell d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b. \end{aligned}$$

Under the null hypothesis (2.11), I compute the mean (up to the leading term) as follows:

$$\begin{aligned} E(Q) &= \frac{1}{T^2} \iint_{(0,T]^2} \int_0^T 1_{\{\ell \in \cap_{i=1}^2 [t_i-T, t_i]\}} \left[\frac{1}{B}w\left(\frac{\ell}{B}\right) + o(1) \right] E\left(d\hat{\epsilon}_{t_1-\ell}^a d\hat{\epsilon}_{t_2-\ell}^a\right) d\ell E\left(d\hat{\epsilon}_{t_1}^b d\hat{\epsilon}_{t_2}^b\right) \\ &= \frac{\lambda^a \lambda^b}{T^2} \int_0^T \int_{t-T}^t \frac{1}{B}w\left(\frac{\ell}{B}\right) d\ell dt. \end{aligned}$$

By Fubini's theorem, the last line becomes

$$E(Q) = \frac{\lambda^a \lambda^b}{T} \int_0^T \frac{1}{B}w\left(\frac{\ell}{B}\right) \left(1 - \frac{\ell}{T}\right) d\ell,$$

so that $E(Q^H) = C^H + o(1)$. By similar techniques as I obtained (A.7), I compute the variance (up to the leading term) as follows:

$$\text{Var}(Q) = 2 \frac{(\lambda^a \lambda^b)^2}{T^2} \int_0^T \frac{1}{B^2} w^2\left(\frac{\ell}{B}\right) \left(1 - \frac{\ell}{T}\right)^2 d\ell,$$

so that $\text{Var}(Q^H) = 2D^H + o(1)$.

A.13 Summary of Jarrow and Yu (2001) Model

Suppose that there are two parties (e.g. firms), a and b , whose assets are subject to the risk of default. Apart from its own idiosyncratic risk, the probability of default of each party depends on the default status of the other party. The distribution of τ^k ($k = a, b$), the time to default by party k , can be fully characterized by the *conditional intensity function*, $\lambda^k(t|\mathcal{F}_{t-}) = \lim_{\Delta t \downarrow 0} (\Delta t)^{-1} P(\tau^k \in [t, t + \Delta t] | \mathcal{F}_{t-})$ where $\mathcal{F} = (\mathcal{F}_t)$ is the natural filtration generated by the processes $1\{\tau^a \leq t\}$ and $1\{\tau^b \leq t\}$, i.e. $\mathcal{F}_t = \sigma\{1\{\tau^a \leq t\}, 1\{\tau^b \leq t\}\}$. Intuitively, it is the conditional probability that party k will default at time t given the history of the default status of both parties. A simple reduced form counterparty risk model is given as follows:

$$\begin{aligned} \text{for party } a: \quad \lambda^a(t|\mathcal{F}_{t-}) &= \mu^a + \alpha^{ab} 1_{\{\tau^b \leq t\}} \text{ for } t \leq \tau^a, \\ \text{for party } b: \quad \lambda^b(t|\mathcal{F}_{t-}) &= \mu^b + \alpha^{ba} 1_{\{\tau^a \leq t\}} \text{ for } t \leq \tau^b. \end{aligned}$$

This is probably the simplest bivariate default risk model with counterparty risk features represented by the parameters α^{ab} and α^{ba} . For instance, if α^{ab} is positive, then the default by party b increases the chance of default by party a , thus suggesting the existence of counterparty risk from party b to party a .

The above counterparty risk model involving two parties can be readily extended to one involving two portfolios, a and b . (e.g. two industries of firms). Each portfolio contains a large number of homogeneous parties whose individual conditional intensities of defaults take the same piecewise constant form. For $k = a, b$, let τ_i^k be the time of the i^{th} default in portfolio k , and define $N_t^k = \sum_{i=1}^{\infty} 1\{\tau_i^k \leq t\}$ which counts the number of default events in portfolio k up to time t . Now, denote the natural filtration of (N^a, N^b) by $\mathcal{F} = (\mathcal{F}_t)$ where $\mathcal{F}_t = \sigma\{(N_s^a, N_s^b) : s \leq t\}$, and the conditional intensity of default in portfolio

k at time t by $\lambda^k(t|\mathcal{F}_{t-}) = \lim_{\Delta t \downarrow 0} (\Delta t)^{-1} P(N_{t+\Delta t}^k - N_t^k > 0 | \mathcal{F}_{t-})$. Analogous to the counterparty risk model with two parties, a counterparty risk model with two portfolios a and b is defined as follows:

$$\text{for portfolio } a: \lambda^a(t|\mathcal{F}_{t-}) = \mu^a + \alpha^{aa} \sum_{q=1}^{\infty} 1_{\{\tau_q^a \leq t\}} + \alpha^{ab} \sum_{j=1}^{\infty} 1_{\{\tau_j^b \leq t\}}, \quad (\text{A.18})$$

$$\text{for portfolio } b: \lambda^b(t|\mathcal{F}_{t-}) = \mu^b + \alpha^{ba} \sum_{i=1}^{\infty} 1_{\{\tau_i^a \leq t\}} + \alpha^{bb} \sum_{q=1}^{\infty} 1_{\{\tau_q^b \leq t\}}. \quad (\text{A.19})$$

We can rewrite (A.18) and (A.19) in terms of the counting processes N_t^k :

$$\lambda^a(t|\mathcal{F}_{t-}) = \mu^a + \alpha^{aa} N_t^a + \alpha^{ab} N_t^b \text{ for } t \leq \tau_i^a,$$

$$\lambda^b(t|\mathcal{F}_{t-}) = \mu^b + \alpha^{ba} N_t^a + \alpha^{bb} N_t^b \text{ for } t \leq \tau_j^b.$$

With an additional exponential function (or other discount factors) to dampen the feedback effect of each earlier default event, the system of conditional intensities constitutes an *bivariate exponential* (or *generalized*) *Hawkes model* for (N^a, N^b) :

$$\begin{aligned} \lambda^a(t|\mathcal{F}_{t-}) &= \mu^a + \alpha^{aa} \sum_{i=1}^{n^a} 1_{\{\tau_i^a \leq t\}} e^{-\beta^{aa}(t-\tau_i^a)} + \alpha^{ab} \sum_{j=1}^{n^b} 1_{\{\tau_j^b \leq t\}} e^{-\beta^{ab}(t-\tau_j^b)} \\ &= \mu^a + \alpha^{aa} \int_0^t e^{-\beta^{aa}(t-s)} dN_s^a + \alpha^{ab} \int_0^t e^{-\beta^{ab}(t-u)} dN_u^b, \\ \lambda^b(t|\mathcal{F}_{t-}) &= \mu^b + \alpha^{ba} \sum_{i=1}^{n^a} 1_{\{\tau_i^a \leq t\}} e^{-\beta^{ba}(t-\tau_i^a)} + \alpha^{bb} \sum_{j=1}^{n^b} 1_{\{\tau_j^b \leq t\}} e^{-\beta^{bb}(t-\tau_j^b)} \\ &= \mu^b + \alpha^{ba} \int_0^t e^{-\beta^{ba}(t-s)} dN_s^a + \alpha^{bb} \int_0^t e^{-\beta^{bb}(t-u)} dN_u^b. \end{aligned}$$

To test for the existence of Granger causality based on this model, we can estimate the parameters α^{ab} and α^{ba} and test if they are significant. However, this parametric bivariate model is only one of the many possible ways that the

conditional intensities of default from two portfolios can interact with one another. The nonparametric test in this chapter can detect Granger causality without making a strong parametric assumption on the bivariate point process of defaults.

A.14 Q Test for Non-stationary Point Processes

There is room for relaxing the stationary assumption on $\mathbf{N} = (N^a, N^b)$, as the asymptotic theory of the test statistic Q relies on the stationarity of only the cross covariance of the innovations $d\epsilon_t^a$ and $d\epsilon_t^b$ from the two marginal processes. First order stationarity of each marginal point process was imposed to merely facilitate (unconditional) moment calculations of Q , and was not necessary if one has a model for the time-varying marginal intensities. To this end, assumption (A2) is replaced by the following weaker assumption:

Assumption (A2b) The cross covariance density function of the innovations $d\epsilon_t^a$ and $d\epsilon_t^b$ from the two marginal processes is a function of the time difference only, i.e. $E(d\epsilon_s^a d\epsilon_t^b) = \gamma(t - s) ds dt$.

Recall that $\gamma(\cdot)$ is the *reduced form cross covariance density function* of the innovations $d\epsilon_t^a$ and $d\epsilon_t^b$, which is well-defined under Assumption (A2b).

Assumption (A2b) encompasses the case that N^k is non-stationary in mean, i.e. $E(dN_t^k)$ is time-varying. Suppose the conditional intensity λ_t^k is specified by a deterministic parametric model, so that $\lambda_t^k \equiv E(dN_t^k | \mathcal{F}_t^k) \equiv \lambda^k(t; \theta^k)$. By iterated expectations, we have $E(dN_t^k) = E(\lambda^k(t; \theta^k)) = \lambda^k(t; \theta^k) = \lambda_t^k$.

Theorem 29 Suppose that the conditional intensities λ_t^k ($k = a, b$) are deterministic. Under assumptions (A1,2b,3,4a,5,6) and the null hypothesis (2.11), the normalized test statistic

$$J = \frac{\tilde{Q} - E(\tilde{Q})}{\sqrt{\text{Var}(\tilde{Q})}} \quad (\text{A.20})$$

converges in distribution to a standard normal random variable as $T \rightarrow \infty$, $H \rightarrow \infty$ and $H/T \rightarrow 0$, where the mean and variance of \tilde{Q} are given as follows:

$$\begin{aligned} E(\tilde{Q}) &= \frac{1}{TH} \kappa_2 \int_I w_B(\ell) \left(\frac{1}{T} \int_{0 \vee \ell}^{T \wedge (T+\ell)} \lambda_{t-\ell}^a \lambda_t^b dt \right) d\ell + o\left(\frac{1}{TH}\right), \\ \text{Var}(\tilde{Q}) &= \frac{2}{T^2 H} \kappa_4 \int_I w_B^2(\ell) \left(\frac{1}{T} \int_{0 \vee \ell}^{T \wedge (T+\ell)} \lambda_{t-\ell}^a \lambda_t^b dt \right)^2 d\ell + o\left(\frac{1}{T^2 H}\right). \end{aligned}$$

Proof. It suffices to verify the mean and the variance of Q as the asymptotic theory is the same as that in the second-order stationary case. Now, the mean of Q is

$$\begin{aligned} E(Q) &= \frac{1}{T^2 H^2} \iint_{(0,T]^2} \int_I w_B(\ell) K^2\left(\frac{t-s-\ell}{H}\right) d\ell \lambda_s^a \lambda_t^b ds dt \\ &= \frac{T^3}{T^2 H^2} \iint_{(0,1]^2} \int_{I/T} w_B(T\sigma) K^2\left(\frac{v-u-\sigma}{H/T}\right) d\sigma \lambda_{Tu}^a \lambda_{Tv}^b du dv \\ &= \frac{1}{T^2 h^2} \iint_{(0,1]^2} \int_{I/T} w_b(\sigma) K^2\left(\frac{v-u-\sigma}{h}\right) d\sigma \lambda_{Tu}^a \lambda_{Tv}^b du dv. \end{aligned}$$

As $h \rightarrow 0$,

$$\begin{aligned} &\iint_{(0,1]^2} \frac{1}{h^2} K^2\left(\frac{v-u-\sigma}{h}\right) \lambda_{Tu}^a \lambda_{Tv}^b du dv \\ &= \frac{1}{h} \int_0^1 \int_{(v-1-\sigma)/h}^{(v-\sigma)/h} K^2(x) \lambda_{T(v-\sigma-xh)}^a \lambda_{Tv}^b dx dv \\ &= \frac{1}{h} \int_0^1 1_{\{0 < v-\sigma < 1\} \cap [0,1]} \lambda_{T(v-\sigma)}^a \lambda_{Tv}^b dv \int_{-\infty}^{\infty} K^2(x) dx + o\left(\frac{1}{h}\right) \\ &= \frac{1}{h} \kappa_2 \int_{0 \vee \sigma}^{1 \wedge (1+\sigma)} \lambda_{T(v-\sigma)}^a \lambda_{Tv}^b dv + o\left(\frac{1}{h}\right), \end{aligned}$$

where $\kappa_2 = \int_{-\infty}^{\infty} K^2(x)dx$ (from assumption (A4a)). As $T \rightarrow \infty$, $Th = H \rightarrow \infty$ and $h = H/T \rightarrow 0$, the asymptotic mean of Q under the null hypothesis is given by

$$\begin{aligned} E(Q) &= \frac{1}{T^2 h} \kappa_2 \int_{I/T} w_b(\sigma) \int_{0 \vee \sigma}^{1 \wedge (1+\sigma)} \lambda_{T(v-\sigma)}^a \lambda_{Tv}^b dv d\sigma + o\left(\frac{1}{T^2 h}\right) \\ &= \frac{1}{T^2 h} \kappa_2 \int_I \frac{1}{b} w\left(\frac{\ell}{bT}\right) \left(\frac{1}{T} \int_{0 \vee \ell}^{T \wedge (T+\ell)} \lambda_{t-\ell}^a \lambda_t^b dt \right) \frac{1}{T} d\ell + o\left(\frac{1}{T^2 h}\right) \\ &= \frac{1}{TH} \kappa_2 \int_I w_B(\ell) \left(\frac{1}{T} \int_{0 \vee \ell}^{T \wedge (T+\ell)} \lambda_{t-\ell}^a \lambda_t^b dt \right) d\ell + o\left(\frac{1}{TH}\right). \end{aligned}$$

On the other hand, the dominating term of the variance is

$$\begin{aligned} Var(Q) &= \frac{2}{(TH)^4} \iint_{I^2} w_B(\ell_1) w_B(\ell_2) \iiint_{(0,T]^4} K\left(\frac{t_1-s_1-\ell_1}{H}\right) K\left(\frac{t_2-s_2-\ell_2}{H}\right) \\ &\quad K\left(\frac{t_1-s_1-\ell_2}{H}\right) K\left(\frac{t_2-s_2-\ell_1}{H}\right) \lambda_{s_1}^a \lambda_{s_2}^a \lambda_{t_1}^b \lambda_{t_2}^b ds_1 ds_2 dt_1 dt_2 d\ell_1 d\ell_2 \\ &= \frac{2T^4}{(TH)^4} \iint_{(I/T)^2} \frac{T^2}{B^2} w\left(\frac{\sigma_1}{B/T}\right) w\left(\frac{\sigma_2}{B/T}\right) \iint_{(0,1]^2} \int_{v_2-1}^{v_2} \int_{v_1-1}^{v_1} K\left(\frac{u-\sigma_1}{H/T}\right) K\left(\frac{v-\sigma_1}{H/T}\right) \\ &\quad K\left(\frac{u-\sigma_2}{H/T}\right) K\left(\frac{v-\sigma_2}{H/T}\right) \lambda_{T(v_1-u)}^a \lambda_{T(v_2-u)}^a \lambda_{Tv_1}^b \lambda_{Tv_2}^b du dv dv_1 dv_2 d\sigma_1 d\sigma_2 \\ &= \frac{2}{T^4 h} \int_{I/T} w_b(\sigma_1) w_b(\sigma_1 - zh) \iint_{(0,1]^2} \int_{\frac{\sigma_1-c_1/T}{h}}^{\frac{\sigma_1-c_1/T}{h}} \int_{\frac{v_2-1-\sigma_1}{h}}^{\frac{v_2-\sigma_1}{h}} \int_{\frac{v_1-1-\sigma_1}{h}}^{\frac{v_1-\sigma_1}{h}} K(x) K(y) \\ &\quad K(x+z) K(y+z) \lambda_{T(v_1-\sigma_1+xh)}^a \lambda_{T(v_2-\sigma_1+yh)}^a \lambda_{Tv_1}^b \lambda_{Tv_2}^b dx dy dz dv_1 dv_2 d\sigma_1 \\ &= \frac{2}{T^4 h} \int_{I/T} w_b(\sigma_1) [w_b(\sigma_1) - zh w_b'(\sigma_1)] \iiint_{\mathbb{R}^3} K(x) K(y) K(x+z) K(y+z) dx dy dz \\ &\quad \iint_{(0,1]^2} 1_{\{(v_1, v_2): 0 \vee \sigma_1 < v_i < (1+\sigma_1) \wedge 1\}} \lambda_{T(v_1-\sigma_1)}^a \lambda_{T(v_2-\sigma_1)}^a \lambda_{Tv_1}^b \lambda_{Tv_2}^b dv_1 dv_2 d\sigma_1 + o\left(\frac{1}{T^3 h}\right) \\ &= \frac{2}{T^2 H} \kappa_4 \int_I w_B^2(\ell) \left(\frac{1}{T} \int_{0 \vee \ell}^{T \wedge (T+\ell)} \lambda_{t-\ell}^a \lambda_t^b dt \right)^2 d\ell + o\left(\frac{1}{T^2 H}\right). \end{aligned}$$

■

APPENDIX B
CHAPTER 3 APPENDIX

periods	Jul90—Mar91		Mar01—Nov02		Mar01—Nov01		Dec07—Jun09	
	$\widehat{\theta}$	t stat	$\widehat{\theta}$	t stat	$\widehat{\theta}$	t stat	$\widehat{\theta}$	t stat
ω^a	-7.65	-72.14	-8.37	-25.18	-9.54	-16.80	-9.32	-39.78
ω^b	-9.14	-27.15	-9.53	-25.43	-10.15	-17.35	-10.71	-27.27
β^{aa}	0.83	6.53	0.04	0.96	0.66	3.93	0.25	7.12
β^{ab}	0.10	6.37	1.72	7.82	1.91	2.64	0.82	38.91
β^{ba}	1.11	1.82	0.02	0.99	0.08	2.76	0.29	8.26
β^{bb}	0.49	2.35	0.92	19.49	0.31	1.97	0.60	23.07
α^{aa}	0.30	0.26	-0.14	-1.43	0.78	10.20	1.32	20.98
α^{ab}	-4.07	-1.73	0.29	5.60	0.51	3.03	0.75	17.26
α^{ba}	1.12	2.23	-120.96	-0.95	-17.84	-1.82	-5.92	-8.46
α^{bb}	4.20	14.92	3.53	7.76	6.38	7.18	4.05	121.17
(n^a, n^b)		(17,11)		(106,56)		(44,32)		(59,42)
L		-99.98		-0.50×10^3		-239.41		-313.46
	stat	p-val	stat	p-val	stat	p-val	stat	p-val
LB_3	1.36	0.243	1.42	0.234	0.75	0.388	10.42	0.001
LB_5	2.10	0.552	5.88	0.118	9.48	0.024	11.26	0.010
LB_{10}	6.75	0.564	6.71	0.568	16.91	0.031	14.46	0.071
LB_{20}	16.00	0.592	10.55	0.912	27.11	0.077	29.59	0.042
$LB1_3$	0.52	0.470	0.31	0.579	3.81	0.051	1.93	0.165
$LB1_5$	1.56	0.670	1.44	0.696	4.58	0.206	2.41	0.492
$LB1_{10}$	2.49	0.962	4.54	0.806	10.97	0.204	3.48	0.900
$LB1_{20}$	-	-	10.86	0.900	21.45	0.257	11.43	0.875
$LB2_3$	0.65	0.422	0.02	0.889	0.01	0.933	0.44	0.507
$LB2_5$	1.41	0.703	0.02	0.999	0.01	1.000	0.45	0.929
$LB2_{10}$	-	-	0.04	1.000	0.03	1.000	0.59	1.000
$LB2_{20}$	-	-	0.08	1.000	0.04	1.000	0.90	1.000

Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b (sample sizes n^a and n^b) corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The sampling windows are measured in days around the reference point. The estimates of the parameters in the ACI(1,1) model were presented along with the t statistics. L is the log-likelihood value. LB_M , $LB1_M$ and $LB2_M$ are the statistic values of the Ljung-Box test (with maximum lag M) on the residuals of the pooled point process, the first and the second marginal point processes, respectively.

Table B.1: Estimated bivariate ACI(1,1) models on point process data of bankruptcy filings over recessions and crises.

periods	Apr91-Feb01		Dec01-Nov07		Dec02-Nov07		Jul09-Nov11	
	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat
ω^a	-9.27	-16.13	-8.24	-21.90	-8.68	-21.68	-10.85	-15.99
ω^b	-7.94	-4.85	-9.92	-22.75	-10.37	-18.36	-9.29	-6.54
β^{aa}	0.32	3.10	0.72	13.88	0.66	9.17	0.87	7.58
β^{ab}	0.54	2.60	0.77	6.43	0.77	5.75	0.02	0.15
β^{ba}	0.68	6.26	0.25	4.60	0.29	4.58	1.01	1.58
β^{bb}	0.35	1.75	0.19	3.75	0.23	3.25	0.05	0.25
α^{aa}	2.67	5.28	0.33	4.34	0.55	5.29	2.99	2.83
α^{ab}	-1.87	-1.33	0.22	2.00	0.32	2.99	-25.42	-0.15
α^{ba}	0.27	0.60	-4.41	-3.91	-4.68	-3.31	0.74	0.96
α^{bb}	0.97	0.70	5.27	16.32	5.24	13.43	0.26	1.23
(n^a, n^b)	(151,162)		(160,70)		(105,50)		(36,51)	
L	-1.26×10^3		-0.84×10^3		-0.62×10^3		-339.64	
	stat	p-val	stat	p-val	stat	p-val	stat	p-val
LB_3	3.21	0.073	1.53	0.216	2.32	0.128	3.92	0.048
LB_5	6.72	0.081	3.39	0.336	3.35	0.340	3.93	0.270
LB_{10}	8.14	0.420	5.28	0.727	6.14	0.631	18.15	0.020
LB_{20}	14.27	0.712	24.63	0.135	18.79	0.405	37.92	0.004
$LB1_3$	0.82	0.367	1.83	0.176	2.27	0.132	0.25	0.614
$LB1_5$	0.90	0.826	2.46	0.483	4.52	0.210	0.29	0.962
$LB1_{10}$	1.07	0.998	7.73	0.460	7.94	0.439	0.32	1.000
$LB1_{20}$	1.33	1.000	19.82	0.343	18.28	0.438	0.70	1.000
$LB2_3$	5.32	0.021	0.02	0.892	0.02	0.876	6.47	0.011
$LB2_5$	6.05	0.109	0.02	0.999	0.03	0.999	9.19	0.027
$LB2_{10}$	8.13	0.421	0.03	1.000	0.04	1.000	16.88	0.031
$LB2_{20}$	19.42	0.367	0.03	1.000	0.05	1.000	19.76	0.346

Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b (sample sizes n^a and n^b) corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The sampling windows are measured in days around the reference point. The estimates of the parameters in the ACI(1,1) model were presented along with the t statistics. L is the log-likelihood value. LB_M , $LB1_M$ and $LB2_M$ are the statistic values of the Ljung-Box test (with maximum lag M) on the residuals of the pooled point process, the first and the second marginal point processes, respectively.

Table B.2: Estimated bivariate ACI(1,1) models on point process data of bankruptcy filings over non-recession periods.

	quarterly		monthly		biweekly		weekly	
	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat
ω^a	-0.14	-0.05	0.39	0.43	0.57	2.17	0.43	3.47
ω^b	1.65	1.87	1.20	2.12	0.74	3.01	0.38	3.73
β^{aa}	-0.78	-2.05	0.42	1.97	0.68	4.94	0.37	3.42
β^{ab}	2.95	3.78	0.72	1.66	-0.12	-0.69	-0.00	-0.02
β^{ba}	-0.28	-2.30	0.21	1.57	0.28	2.19	0.19	2.17
β^{bb}	1.37	5.38	0.18	0.66	-0.13	-0.78	-0.03	-0.30
ρ^{ab}	-0.74		0.21		0.26		0.25	
σ^{aa}	13.76		4.67		1.23		0.69	
σ^{ab}	-3.34		0.61		0.30		0.14	
σ^{bb}	1.46		1.85		1.07		0.46	
	stat	p-val	stat	p-val	stat	p-val	stat	p-val
H_3	13.54	0.09	9.67	0.29	11.24	0.19	23.74	0.00
JB^a	0.62	0.74	7.22	0.03	0.73	0.70	34.68	0.00
JB^b	0.77	0.68	1.09	0.58	15.90	0.00	10.54	0.01
DW^a	1.75		2.19		2.37		2.30	
DW^b	2.54		2.20		2.06		2.12	
n	7		19		42		84	

Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The estimates of the parameters in the VAR(1) model were presented along with the t statistics. $\rho^{ab} = \sigma^{ab} / \sqrt{\sigma^{aa}\sigma^{bb}}$. n = sample length of the count sequences. H_3 = Hosking's (1980) multivariate portmanteau test on residuals. JB^k = univariate Jarque-Bera normality test on group k residuals. DW^k = univariate Durbin-Watson test on group k residuals.

Table B.3: Estimated VAR(1) model on the time series of bankruptcy counts over Dec 2007 - June 2009 with varying sampling frequencies.

	quarterly		monthly		biweekly		weekly	
	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat
ω^a	10.46	1.92	1.97	1.51	1.79	3.96	1.18	6.32
ω^b	5.95	2.97	2.10	2.13	1.29	4.12	0.59	4.36
β^{aa}	-0.42	-1.11	0.30	1.46	0.02	0.14	0.01	0.12
β^{ab}	1.44	1.90	0.63	2.05	0.43	1.99	-0.04	-0.26
β^{ba}	-0.39	-2.80	-0.05	-0.33	-0.12	-1.17	-0.04	-0.55
β^{bb}	0.93	3.36	0.24	1.02	0.12	0.81	0.07	0.63
ρ^{ab}	-0.30		0.07		0.09		0.05	
σ^{aa}	30.09		4.92		2.42		1.24	
σ^{ab}	-3.31		0.27		0.16		0.04	
σ^{bb}	4.06		2.82		1.16		0.66	
	stat	p-val	stat	p-val	stat	p-val	stat	p-val
H_3	15.79	0.05	7.27	0.51	8.78	0.36	14.25	0.08
JB^a	0.93	0.63	0.98	0.61	1.16	0.56	11.94	0.00
JB^b	0.71	0.70	0.65	0.72	3.73	0.15	30.01	0.00
DW^a	2.03		1.49		1.98		1.99	
DW^b	1.14		2.16		1.89		1.82	
n	8		21		47		92	

Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The estimates of the parameters in the VAR(1) model were presented along with the t statistics. $\rho^{ab} = \sigma^{ab} / \sqrt{\sigma^{aa}\sigma^{bb}}$. n = sample length of the count sequences. H_3 = Hosking's (1980) multivariate portmanteau test on residuals. JB^k = univariate Jarque-Bera normality test on group k residuals. DW^k = univariate Durbin-Watson test on group k residuals.

Table B.4: Estimated VAR(1) model on the time series of bankruptcy counts over March 2001 - November 2002 with varying sampling frequencies.

	monthly		biweekly		weekly	
	$\widehat{\theta}$	t stat	$\widehat{\theta}$	t stat	$\widehat{\theta}$	t stat
ω^a	1.00	0.89	0.50	1.65	0.30	2.45
ω^b	1.26	1.62	0.60	1.99	0.27	2.16
β^{aa}	0.38	0.79	0.22	0.94	0.01	0.06
β^{ab}	0.13	0.17	0.29	1.16	0.39	2.40
β^{ba}	0.18	0.55	0.10	0.44	0.05	0.33
β^{bb}	-0.16	-0.32	-0.18	-0.72	-0.01	-0.08
ρ^{ab}	0.51		0.15		0.11	
σ^{aa}	3.10		0.89		0.37	
σ^{ab}	1.10		0.13		0.04	
σ^{bb}	1.49		0.88		0.38	
	stat	p-val	stat	p-val	stat	p-val
H_3	7.22	0.51	11.30	0.19	2.72	0.95
JB^a	1.73	0.42	1.35	0.51	7.29	0.03
JB^b	0.79	0.67	3.82	0.15	36.89	0.00
DW^a	1.63		1.98		1.94	
DW^b	1.92		2.09		2.00	
n	9		20		40	

Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The estimates of the parameters in the VAR(1) model were presented along with the t statistics. $\rho^{ab} = \sigma^{ab} / \sqrt{\sigma^{aa}\sigma^{bb}}$. n = sample length of the count sequences. H_3 = Hosking's (1980) multivariate portmanteau test on residuals. JB^k = univariate Jarque-Bera normality test on group k residuals. DW^k = univariate Durbin-Watson test on group k residuals.

Table B.5: Estimated VAR(1) model on the time series of bankruptcy counts over July 1990 - March 1991 with varying sampling frequencies.

	quarterly		monthly		biweekly		weekly	
	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat
ω^a	1.10	1.84	0.59	3.41	0.39	5.55	0.24	8.01
ω^b	1.20	1.44	0.57	3.02	0.41	5.44	0.24	7.29
β^{aa}	0.59	3.88	0.33	3.77	0.12	1.97	0.13	3.03
β^{ab}	0.12	1.08	0.18	2.27	0.18	3.14	0.03	0.72
β^{ba}	0.61	2.87	0.47	4.95	0.16	2.37	0.14	2.85
β^{bb}	0.15	0.91	0.14	1.71	0.20	3.26	0.10	2.39
ρ^{ab}	0.25		0.05		0.04		0.02	
σ^{aa}	4.37		1.51		0.69		0.31	
σ^{ab}	1.51		0.08		0.03		0.01	
σ^{bb}	8.55		1.81		0.79		0.38	
	stat	p-val	stat	p-val	stat	p-val	stat	p-val
H_3	16.86	0.03	8.80	0.36	35.89	0.00	27.53	0.00
JB^a	1.62	0.44	5.12	0.08	411.95	0.00	675.32	0.00
JB^b	18.74	0.00	55.88	0.00	292.46	0.00	885.48	0.00
DW^a	2.49		2.03		2.07		2.01	
DW^b	2.05		2.04		2.13		2.04	
n	40		119		259		518	

Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The estimates of the parameters in the VAR(1) model were presented along with the t statistics. $\rho^{ab} = \sigma^{ab} / \sqrt{\sigma^{aa}\sigma^{bb}}$. n = sample length of the count sequences. H_3 = Hosking's (1980) multivariate portmanteau test on residuals. JB^k = univariate Jarque-Bera normality test on group k residuals. DW^k = univariate Durbin-Watson test on group k residuals.

Table B.6: Estimated VAR(1) model on the time series of bankruptcy counts over April 1990 - March 1991 with varying sampling frequencies.

	quarterly		monthly		biweekly		weekly	
	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat
ω^a	1.11	0.93	0.59	2.06	0.56	4.73	0.31	6.30
ω^b	1.38	1.80	0.38	2.21	0.31	4.10	0.13	4.19
β^{aa}	0.46	2.26	0.36	3.14	0.18	2.13	0.15	2.50
β^{ab}	0.64	1.59	0.55	2.58	0.16	1.19	0.06	0.62
β^{ba}	0.02	0.19	0.15	2.16	0.07	1.24	0.05	1.31
β^{bb}	0.30	1.14	0.13	1.01	-0.06	-0.64	0.11	1.76
ρ^{ab}	0.39		0.06		0.09		0.01	
σ^{aa}	9.45		2.13		1.06		0.42	
σ^{ab}	2.39		0.08		0.06		0.00	
σ^{bb}	3.94		0.78		0.45		0.18	
	stat	p-val	stat	p-val	stat	p-val	stat	p-val
H_3	5.50	0.70	5.84	0.67	14.34	0.07	18.98	0.02
JB^a	0.66	0.72	6.42	0.04	232.15	0.00	271.27	0.00
JB^b	2.26	0.32	47.64	0.00	206.82	0.00	530.21	0.00
DW^a	2.08		2.12		1.98		2.14	
DW^b	1.69		2.01		1.93		1.99	
n	21		60		131		262	

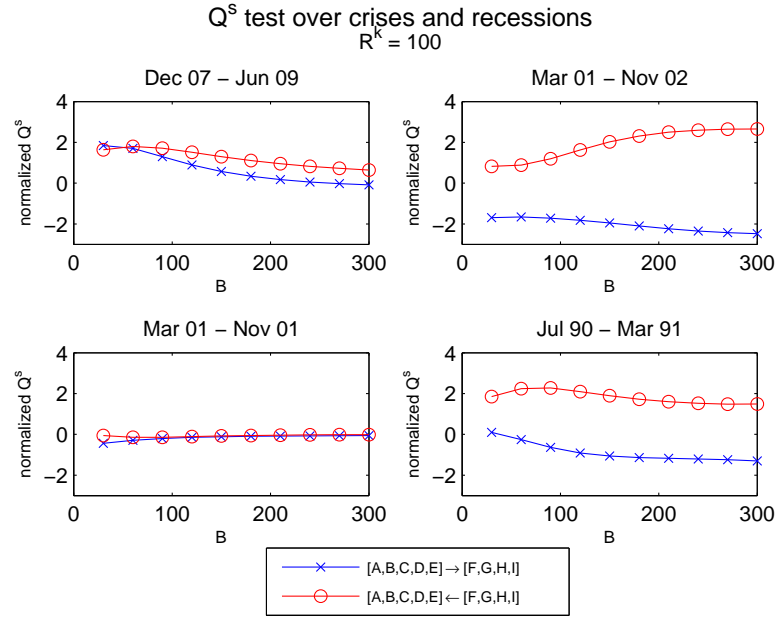
Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The estimates of the parameters in the VAR(1) model were presented along with the t statistics. $\rho^{ab} = \sigma^{ab} / \sqrt{\sigma^{aa}\sigma^{bb}}$. n = sample length of the count sequences. H_3 = Hosking's (1980) multivariate portmanteau test on residuals. JB^k = univariate Jarque-Bera normality test on group k residuals. DW^k = univariate Durbin-Watson test on group k residuals.

Table B.7: Estimated VAR(1) model on the time series of bankruptcy counts over December 2002 - November, 2007 with varying sampling frequencies.

	quarterly		monthly		biweekly		weekly	
	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat	$\hat{\theta}$	t stat
ω^a	-0.29	-0.27	0.22	0.90	0.13	1.49	0.13	3.14
ω^b	0.20	0.18	0.22	0.91	0.25	2.06	0.16	2.80
β^{aa}	0.17	0.40	0.58	2.95	0.21	1.78	0.21	2.25
β^{ab}	0.57	1.46	0.04	0.28	0.32	3.48	0.04	0.61
β^{ba}	0.44	1.01	0.92	4.61	0.57	3.54	0.36	2.78
β^{bb}	0.27	0.67	-0.05	-0.30	-0.00	-0.02	0.02	0.21
ρ^{ab}	0.66		0.26		0.30		0.32	
σ^{aa}	6.16		1.08		0.36		0.17	
σ^{ab}	4.18		0.29		0.15		0.07	
σ^{bb}	6.51		1.12		0.66		0.32	
	stat	p-val	stat	p-val	stat	p-val	stat	p-val
H_3	16.65	0.03	9.97	0.27	17.08	0.03	33.29	0.00
JB^a	0.48	0.79	116.64	0.00	63.06	0.00	159.84	0.00
JB^b	0.37	0.83	11.24	0.00	69.44	0.00	246.14	0.00
DW^a	1.53		2.27		1.93		2.06	
DW^b	3.18		1.91		2.34		2.17	
n	10		29		64		127	

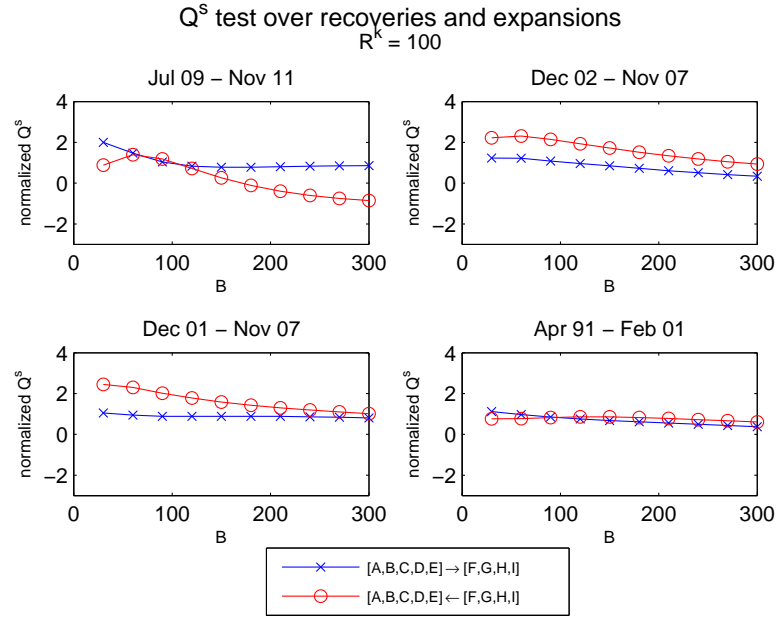
Bankruptcies are classified into two categories (upper and lower ends of a typical supply chain) according to the SIC divisions of the filing companies, with groups a and b corresponding to SIC divisions $\{A,B,C,D,E\}$ and $\{F,G,H,I\}$, respectively. The estimates of the parameters in the VAR(1) model were presented along with the t statistics. $\rho^{ab} = \sigma^{ab} / \sqrt{\sigma^{aa}\sigma^{bb}}$. n = sample length of the count sequences. H_3 = Hosking's (1980) multivariate portmanteau test on residuals. JB^k = univariate Jarque-Bera normality test on group k residuals. DW^k = univariate Durbin-Watson test on group k residuals.

Table B.8: Estimated VAR(1) model on the time series of bankruptcy counts over July, 2009 - November, 2011 with varying sampling frequencies.



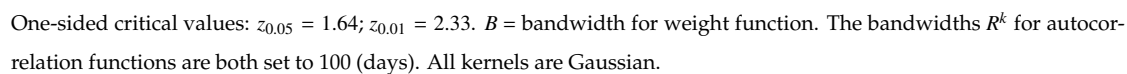
One-sided critical values: $z_{0.05} = 1.64$; $z_{0.01} = 2.33$. B = bandwidth for weight function. The bandwidths R^k for autocorrelation functions are both set to 100 (days). All kernels are Gaussian.

Figure B.1: Q^s test between upstream and downstream of a supply chain during crises and recessions.

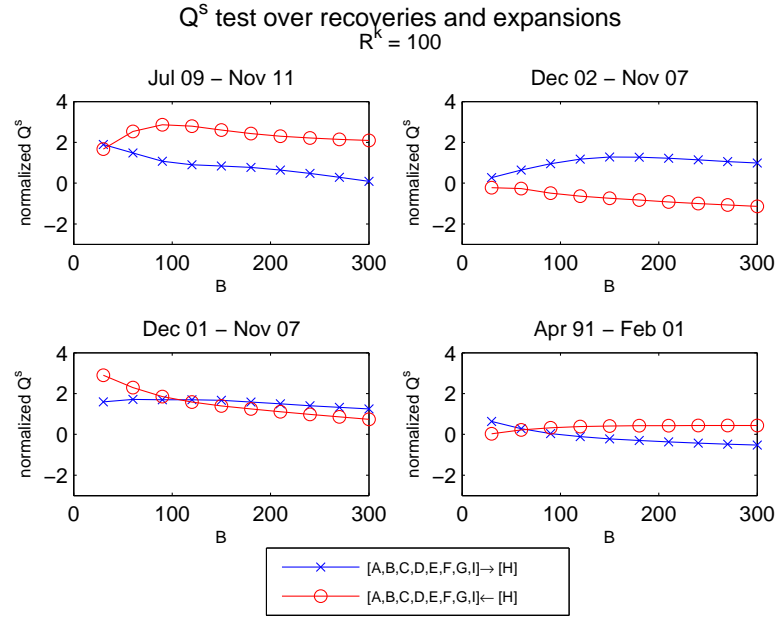


One-sided critical values: $z_{0.05} = 1.64$; $z_{0.01} = 2.33$. B = bandwidth for weight function. The bandwidths R^k for autocorrelation functions are both set to 100 (days). All kernels are Gaussian.

Figure B.2: Q^s test between upstream and downstream of a supply chain during non-recession periods.



186



One-sided critical values: $z_{0.05} = 1.64$; $z_{0.01} = 2.33$. B = bandwidth for weight function. The bandwidths R^k for autocorrelation functions are both set to 100 (days). All kernels are Gaussian.

Figure B.4: Q^s test between Wall Street and Main Street during non-recession periods.

APPENDIX C
CHAPTER 4 APPENDIX

C.1 Proof of Theorem 19

Recall that at the true parameter value θ_0 , the generalized innovations $\varepsilon_i = 1 - \int_{T_{i-1}}^{T_i} h(t; \theta_0) dt$ are iid with mean 0. Let the vector of M autocovariances of the generalized innovations be $\mathbf{r} = (r_1, r_2, \dots, r_M)'$, where $r_m = \frac{1}{n} \sum_{i=m+1}^n \varepsilon_i \varepsilon_{i-m}$. Let $\mathbf{X} = -\mathbb{E}[\partial \mathbf{r} / \partial \theta]$ be an $(M \times D)$ matrix. Its (m, d) entry is then given by

$$\begin{aligned} x_{mj} &= -\frac{1}{n} \mathbb{E} \left(\frac{\partial r_m}{\partial \theta_j} \right) \\ &= -\frac{1}{n} \sum_{i=m+1}^n \left[\mathbb{E} \left(\varepsilon_i \frac{\partial \varepsilon_{i-m}}{\partial \theta_j} \right) + \mathbb{E} \left(\varepsilon_{i-m} \frac{\partial \varepsilon_i}{\partial \theta_j} \right) \right] \end{aligned} \quad (\text{C.1})$$

but by iterated expectations, the first expectation is equal to zero, while the second expectation can be simplified as:

$$\begin{aligned} \mathbb{E} \left(\varepsilon_{i-m} \frac{\partial \varepsilon_i}{\partial \theta_j} \right) &= \mathbb{E} \left(-\varepsilon_{i-m} \int_{T_{i-1}}^{T_i} \frac{\partial}{\partial \theta_j} h(t; \theta_0) dt \right) \\ &= -\mathbb{E} \left[\varepsilon_{i-m} \int_{T_{i-1}}^{T_i} \left(\frac{\partial}{\partial \theta_j} \log h(t; \theta_0) \right) h(t; \theta_0) dt \right] \\ &= -\mathbb{E} \left[\varepsilon_{i-m} \frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \right], \end{aligned}$$

where the first equality follows from Leibniz theorem that allows for interchanging the order of integration and differentiation, and the last equality follows from lemma 30. Therefore, \mathbf{X} is as given in (4.4) in the theorem.

Let $\mathbf{Y} = \mathbb{E}[\partial \ell(\theta_0) / \partial \theta \cdot \mathbf{r}']$ be an $(D \times M)$ matrix. Our goal is to show that

$$\mathbf{X} = \mathbf{Y}'. \quad (\text{C.2})$$

Let the total number of observed times be $n = N(T)$. Recall that the log-likelihood function evaluated at the true parameter is

$$\begin{aligned}\ell(\theta_0) &= -\Lambda(T) + \sum_{i=1}^n \log h(T_i; \theta_0) \\ &= \sum_{i=1}^n \left[-\int_{T_{i-1}}^{T_i} h(t; \theta_0) dt + \log h(T_i; \theta_0) \right] \\ &= \sum_{i=1}^n [(\varepsilon_i - 1) + \log h(T_i; \theta_0)].\end{aligned}$$

Hence,

$$\begin{aligned}y_{jm} &= \mathbb{E} \left[\frac{\partial \ell(\theta_0)}{\partial \theta_j} r_m \right] = \mathbb{E} \left\{ \sum_{i=1}^n \left[\frac{\partial \varepsilon_i}{\partial \theta_j} + \frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \right] \cdot \frac{1}{n} \sum_{k=m+1}^n \varepsilon_k \varepsilon_{k-m} \right\} \\ &= \frac{1}{n} \mathbb{E} \left\{ \sum_{i=1}^n \left[\frac{\partial \varepsilon_i}{\partial \theta_j} + \frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \right] \sum_{k=m+1}^i \varepsilon_k \varepsilon_{k-m} \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k=m+1}^i \left[\mathbb{E} \left(\frac{\partial \varepsilon_i}{\partial \theta_j} \varepsilon_k \varepsilon_{k-m} \right) + \mathbb{E} \left(\frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \varepsilon_k \varepsilon_{k-m} \right) \right],\end{aligned}$$

where the second equality follows by iterated expectations and by noting that the square bracketed term is measurable with respect to \mathcal{F}_{T_i} . Now, the first expectation becomes:

$$\begin{aligned}&\mathbb{E} \left(\frac{\partial \varepsilon_i}{\partial \theta_j} \varepsilon_k \varepsilon_{k-m} \right) \\ &= \mathbb{E} \left[-\int_{T_{i-1}}^{T_i} \frac{\partial}{\partial \theta_j} h(t; \theta_0) dt \cdot \varepsilon_k \varepsilon_{k-m} \right] \\ &= -\mathbb{E} \left[\int_{T_{i-1}}^{T_i} \left(\frac{\partial}{\partial \theta_j} \log h(t; \theta_0) \right) h(t; \theta_0) dt \cdot \varepsilon_k \varepsilon_{k-m} \right] \\ &= \begin{cases} -\mathbb{E} \left[\left(\frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \right) \cdot \varepsilon_k \varepsilon_{k-m} \right] + \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \right) \cdot \varepsilon_{i-m} \right] & \text{if } k = i; \\ -\mathbb{E} \left[\left(\frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \right) \cdot \varepsilon_k \varepsilon_{k-m} \right] & \text{if } k < i; \end{cases},\end{aligned}$$

where the last line follows from lemma 31 (for $k = i$) and iterative expectations (for $k < i$). Therefore, the (j, m) entry of \mathbf{Y} is given by

$$\begin{aligned}y_{jm} &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_j} \log h(T_i; \theta_0) \right) \cdot \varepsilon_{i-m} \right] \\ &= x_{mj}\end{aligned}$$

and hence $\mathbf{X} = \mathbf{Y}'$. This, together with the result from expanding $\frac{\partial \ell(\hat{\theta})}{\partial \theta}$ around θ_0 :

$$\begin{aligned} 0 &= \frac{\partial \ell(\hat{\theta})}{\partial \theta} = \frac{\partial \ell(\theta_0)}{\partial \theta} + (\hat{\theta} - \theta_0) \frac{\partial^2 \ell(\theta_0)}{\partial \theta \partial \theta'} + o_p\left(\frac{1}{\sqrt{n}}\right) \\ \Rightarrow (\hat{\theta} - \theta_0) &= \mathbf{G}^{-1} \frac{\partial \ell(\theta_0)}{\partial \theta} + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

proves that

$$\begin{aligned} \text{Var}(\sqrt{n}\hat{\mathbf{r}}) &= \text{Var}(\sqrt{n}[\mathbf{r} - \mathbf{X}(\hat{\theta} - \theta_0)]) + o(1) \\ &= \text{Var}(\sqrt{n}\mathbf{r}) + \mathbf{X}\text{Var}[\sqrt{n}(\hat{\theta} - \theta_0)]\mathbf{X}' \\ &\quad - \mathbf{X}\text{Cov}[\sqrt{n}(\hat{\theta} - \theta_0), \mathbf{r}] - \text{Cov}[\sqrt{n}(\hat{\theta} - \theta_0), \mathbf{r}]\mathbf{X}' + o(1) \\ &= \mathbf{I}_M + \mathbf{X}\mathbf{G}^{-1}\mathbf{X}' - \mathbf{X}\mathbf{G}^{-1}\mathbf{Y} - \mathbf{Y}'\mathbf{G}^{-1}\mathbf{X}' + o(1) \\ &= \mathbf{I}_M - \mathbf{X}\mathbf{G}^{-1}\mathbf{X}' + o(1). \\ &= \mathbf{V} + o(1), \end{aligned}$$

as in (4.3).

To show (4.5), we differentiate the log-likelihood function (4.2) twice. The (j, k) entry of the resultant hessian matrix is

$$\frac{\partial^2 \ell(\theta)}{\partial \theta_j \partial \theta_k} = \sum_{i=1}^n \left[- \int_{T_{i-1}}^{T_i} \frac{\partial^2}{\partial \theta_j \partial \theta_k} h(t; \theta) dt + \frac{\partial^2}{\partial \theta_j \partial \theta_k} \log h(T_i; \theta) \right].$$

Note that

$$\frac{\partial}{\partial \theta_j} h(t; \theta) = \left[\frac{\partial}{\partial \theta_j} \log h(t; \theta) \right] h(t; \theta)$$

and

$$\begin{aligned} \frac{\partial^2}{\partial \theta_j \partial \theta_k} h(t; \theta) &= \left[\frac{\partial}{\partial \theta_j} \log h(t; \theta) \right] \left[\frac{\partial}{\partial \theta_k} \log h(t; \theta) \right] h(t; \theta) \\ &\quad + \left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log h(t; \theta) \right] h(t; \theta). \end{aligned}$$

As a result,

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2 \ell(\theta)}{\partial \theta_j \partial \theta_k}\right) &= \sum_{i=1}^n \left\{ -\mathbb{E}\left[\frac{\partial}{\partial \theta_j} \log h(T_i; \theta) \frac{\partial}{\partial \theta_k} \log h(T_i; \theta)\right] \right. \\ &\quad \left. -\mathbb{E}\left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log h(T_i; \theta)\right] + \mathbb{E}\left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log h(T_i; \theta)\right] \right\} \\ &= \sum_{i=1}^n -\mathbb{E}\left[\frac{\partial}{\partial \theta_j} \log h(T_i; \theta) \frac{\partial}{\partial \theta_k} \log h(T_i; \theta)\right].\end{aligned}$$

In getting the first equality above, the integration is simplified by lemma 30 below. Equation (4.5) follows.

Lemma 30 Suppose $h(t)$ is the conditional intensity function associated with the event times $\{T_i\}_{i=0}^\infty$ and $f(t)$ is an $\mathcal{F}_{T_{i-1}}$ -measurable function such that

$$\mathbb{E}\left(\int_{T_{i-1}}^{T_i} |f(t)h(t)| dt \middle| \mathcal{F}_{T_{i-1}}\right) < \infty.$$

Then we have

$$\mathbb{E}\left(\int_{T_{i-1}}^{T_i} f(t)h(t)dt \middle| \mathcal{F}_{T_{i-1}}\right) = \mathbb{E}(f(T_i) | \mathcal{F}_{T_{i-1}}).$$

Proof. See Lemma A1 in the Appendix of Kwok and Li (2008), which proves the special case where $f(t) = f(t - T_{i-1})$. ■

Applying iterated expectations and the above lemma, we conclude that

$$\mathbb{E}\left(\int_{T_{i-1}}^{T_i} f(t)h(t)dt \cdot \varepsilon_{i-m}\right) = \mathbb{E}(f(T_i)\varepsilon_{i-m}).$$

The advantages of this result are twofold: it avoids numerical integrations, and leads to the equality $\mathbf{X} = \mathbf{Y}'$ in (C.2), which simplifies the variance of generalized residual autocovariances to (4.3).

Lemma 31 With the notations in lemma 30, we have

$$\mathbb{E}\left(\int_{T_{i-1}}^{T_i} f(t)h(t)dt \cdot \varepsilon_i \varepsilon_{i-m}\right) = \mathbb{E}(f(T_i)\varepsilon_i \varepsilon_{i-m}) - \mathbb{E}(f(T_i)\varepsilon_{i-m}). \quad (\text{C.3})$$

Proof. This is an application of Lemma 30 and Fubini's theorem. First, we observe that

$$\begin{aligned}
& \mathbb{E}[(1 - \varepsilon_i) 1(T_i > t) | \mathcal{F}_{T_{i-1}}] \\
&= \mathbb{E}\left[\int_{T_{i-1}}^{T_i} h(s) ds \cdot 1(T_i > t) \middle| \mathcal{F}_{T_{i-1}}\right] \\
&= \mathbb{E}\left[\int_{T_{i-1}}^{\infty} h(s) 1(T_i > s) ds \cdot 1(T_i > t) \middle| \mathcal{F}_{T_{i-1}}\right] \\
&= \int_{T_{i-1}}^{\infty} h(s) \mathbb{E}[1(T_i > s \vee t) | \mathcal{F}_{T_{i-1}}] ds
\end{aligned} \tag{C.4}$$

where the last line is obtained by Fubini's theorem and the fact that the conditional intensity function $h(s)$ over $s \in (T_{i-1}, \infty)$ is $\mathcal{F}_{T_{i-1}}$ -measurable (see (C.5) below for the explicit form). Using the notations of Chapter 3 of Brémaud (1981), we denote the conditional distribution and density functions of $T_i - T_{i-1}$ given T_{i-1} by $G^{(i)}(\cdot)$ and $g^{(i)}(\cdot)$, respectively. The conditional intensity function $h(s)$ over $s \in (T_{i-1}, \infty)$ can then be expressed as

$$h(s) = \frac{g^{(i)}(s - T_{i-1})}{1 - G^{(i)}(s - T_{i-1})}. \tag{C.5}$$

Therefore,

$$\mathbb{E}[1(T_i > s \vee t) | \mathcal{F}_{T_{i-1}}] = \begin{cases} 1 - G^{(i)}(t - T_{i-1}) & \text{if } T_{i-1} < s \leq t; \\ 1 - G^{(i)}(s - T_{i-1}) & \text{if } s > t. \end{cases}$$

Continuing from (C.4), we obtain

$$\begin{aligned}
& \mathbb{E}[(1 - \varepsilon_i) 1(T_i > t) | \mathcal{F}_{T_{i-1}}] \\
&= \int_{T_{i-1}}^t h(s) [1 - G^{(i)}(t - T_{i-1})] ds + \int_t^{\infty} h(s) [1 - G^{(i)}(s - T_{i-1})] ds \\
&= [1 - G^{(i)}(t - T_{i-1})] \int_{T_{i-1}}^t h(s) ds + \int_t^{\infty} \frac{g^{(i)}(s - T_{i-1})}{1 - G^{(i)}(s - T_{i-1})} [1 - G^{(i)}(s - T_{i-1})] ds \\
&= [1 - G^{(i)}(t - T_{i-1})] \int_{T_{i-1}}^t h(s) ds + [1 - G^{(i)}(t - T_{i-1})] \\
&= [1 - G^{(i)}(t - T_{i-1})] \left(\int_{T_{i-1}}^t h(s) ds + 1 \right).
\end{aligned}$$

As a result,

$$\begin{aligned}
& \mathbb{E} \left(\int_{T_{i-1}}^{T_i} (1 - \varepsilon_i) f(t) h(t) dt \middle| \mathcal{F}_{T_{i-1}} \right) \\
&= \int_{T_{i-1}}^{\infty} f(t) h(t) \mathbb{E} \left[(1 - \varepsilon_i) 1(T_i > t) \middle| \mathcal{F}_{T_{i-1}} \right] dt \\
&= \int_{T_{i-1}}^{\infty} f(t) \frac{g^{(i)}(t - T_{i-1})}{1 - G^{(i)}(t - T_{i-1})} \left[1 - G^{(i)}(t - T_{i-1}) \right] \left(\int_{T_{i-1}}^t h(s) ds + 1 \right) dt \\
&= \int_{T_{i-1}}^{\infty} f(t) \left(\int_{T_{i-1}}^t h(s) ds + 1 \right) g^{(i)}(t - T_{i-1}) dt \\
&= \mathbb{E} \left[f(T_i) \left(\int_{T_{i-1}}^{T_i} h(s) ds + 1 \right) \middle| \mathcal{F}_{T_{i-1}} \right] \\
&= \mathbb{E} [f(T_i) (2 - \varepsilon_i) | \mathcal{F}_{T_{i-1}}]
\end{aligned}$$

where the second to last equality comes from the definition of $g^{(i)}(\cdot)$, and the last one comes from the definition of generalized innovation $\varepsilon_i = 1 - \int_{T_{i-1}}^{T_i} h(s) ds$.

Lastly, by lemma 30 and iterated expectations,

$$\begin{aligned}
& \mathbb{E} \left(\int_{T_{i-1}}^{T_i} f(t) h(t) dt \cdot \varepsilon_i \varepsilon_{i-m} \right) \\
&= \mathbb{E} \left(\int_{T_{i-1}}^{T_i} f(t) h(t) dt \cdot \varepsilon_{i-m} \right) - \mathbb{E} \left(\int_{T_{i-1}}^{T_i} f(t) h(t) dt \cdot (1 - \varepsilon_i) \varepsilon_{i-m} \right) \\
&= \mathbb{E} (f(T_i) \varepsilon_{i-m}) - \mathbb{E} \left[\mathbb{E} \left(\int_{T_{i-1}}^{T_i} (1 - \varepsilon_i) f(t) h(t) dt \middle| \mathcal{F}_{T_{i-1}} \right) \varepsilon_{i-m} \right] \\
&= \mathbb{E} (f(T_i) \varepsilon_{i-m}) - \mathbb{E} [f(T_i) (2 - \varepsilon_i) \varepsilon_{i-m}] \\
&= \mathbb{E} (f(T_i) \varepsilon_i \varepsilon_{i-m}) - \mathbb{E} (f(T_i) \varepsilon_{i-m}).
\end{aligned}$$

■

BIBLIOGRAPHY

- [1] Aït-Sahalia, Yacine, Julio Cacho-Diaz and Roger J.A. Laeven (2010), Modeling financial contagion using mutually exciting jump processes, Working paper.
- [2] Aït-Sahalia, Yacine and Per A. Mykland (2003), The effects of random and discrete sampling when estimating continuous-time diffusions, *Econometrica* 71, 2, 483–549.
- [3] Acharya, Viral V. and Tanju Yorulmazer (2008), Information contagion and bank herding, *Journal of Money, Credit and Banking* 40, 215–231.
- [4] Allen, Franklin and Douglas Gale (2000), Financial contagion, *Journal of Political Economy* 108, 1, 1–33.
- [5] Battiston, Stefano, Domenico Delli Gatti, Mauro Gallegati, Bruce Greenwald and Joseph E. Stiglitz (2007), Credit chains and bankruptcy propagation in production networks, *Journal of Economic Dynamics and Control* 31, 2061–2084.
- [6] Azizpour, Shahriar, Kay Giesecke and Gustavo Schwenkler (2008), Exploring the sources of default clustering, Working paper.
- [7] Barlett, Maurice Stevenson (1964), The spectral analysis of two-dimensional point processes, *Biometrika*, 51, 3 and 4, 299–311.
- [8] Bowsher, Clive G. (2007), Modelling security market events in continuous time: Intensity based, multivariate point process models, *Journal of Econometrics* 141, 2, 876–912.
- [9] Box, G.E.P. and David A. Pierce (1970), Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association* 65, 332, 1509–1526.
- [10] Brémaud, Pierre (1981), *Point processes and queues: martingale dynamics*, Springer Verlag.
- [11] Brillinger, David R. (1976), Estimation of the second-order intensities of a bivariate stationary point process. *Journal of the Royal Statistical Society Series B* 38, 1, 60–66.

- [12] Chava, Sudheer and Robert Jarrow (2004), Bankruptcy prediction with industry effect, *Review of Finance* 8, 537–569.
- [13] Chiu, Ching Wai, Bjørn Eraker, Andrew T. Foerster, Tae Bong Kim, Hernán D. Seoane (2011), Estimating VAR's sampled at mixed or irregular spaced frequencies : A bayesian approach, Working paper, Economic Research Department, Federal Reserve Bank of Kansas City.
- [14] Christiano, Lawrence J. and Martin Eichenbaum (1987), Temporal aggregation and structural inference in macroeconomics, *Carnegie-Rochester Conference Series on Public Policy* 26, 63–130.
- [15] Comte, Fabienne and Eric Renault (1996), Noncausality in continuous time, *Econometric Theory* 12, 215–256.
- [16] Cox, David R. (1965), On the estimation of the intensity function of a stationary point process. *Journal of the Royal Statistical Society Series B* 27, 2, 332–337.
- [17] Cox, David R. and Peter Adrian Walter Lewis (1972), Multivariate point processes. *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, 2, 401–448.
- [18] Das, Sanjiv, Darrell Duffie, Nikunj Kapadia and Leandro Saita (2007), Common failings: How corporate defaults are correlated, *Journal of Finance* 62, 1, 93–117.
- [19] Daley, Daryl J. and David Vere-Jones (2003), *An introduction to the theory of point processes: elementary theory and methods*, Springer.
- [20] David, Stella Veronica (2008), Central limit theorems for empirical product densities of stationary point processes, Dissertation, Universität Augsburg.
- [21] Diamond, Douglas W. and Robert E. Verrecchia (1987), Constraints on short-selling and asset pricing adjustment to private information, *Journal of Financial Economics* 18, 2, 277–311.
- [22] Diks, Cees and Valentyn Panchenko (2006), A new statistic and practical guidelines for nonparametric Granger causality testing, *Journal of Economic Dynamics and Control* 30, 9-10, 1647–1669.
- [23] Doss, Hani (1989), On estimating the dependence of two point processes, *Annals of Statistics* 17, 2, 749–763.

- [24] Dufour, Alfonso and Robert Engle (2000), Time and the price impact of a trade, *Journal of Finance* 55, 6, 2467–2498.
- [25] Duffie, Darrel, Andreas Eckners, Guillaume Horel, and Leandro Saita (2009), *Frailty correlated default*, *Journal of Finance* 64, 5, 2089–2123.
- [26] Duffie, Darrel and Peter Glynn (2004), Estimation of continuous-time markov processes sampled at random time intervals, *Econometrica* 72, 6, 1773–1808.
- [27] Duffie, Darrell, Leandro Saita and Ke Wang (2007), Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics* 83, 635–665.
- [28] Easley, David and Maureen O’Hara (1987), Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 1, 69–90.
- [29] Easley, David and Maureen O’Hara (1992), Time and the process of security price adjustment, *Journal of Finance* 47, 2, 576–605.
- [30] Engle, Robert F. (2000), The econometrics of ultra-high frequency data, *Econometrica* 68, 1, 1–22.
- [31] Engle, Robert F. and Ta-Chung Liu (1972), Effects of aggregation over time on dynamic characteristics of an economic model, in Burt G. Hickman, ed., *Econometric Models of Cyclical Behavior, Studies in Income and Wealth*, v. 2, no. 36, National Bureau of Economic Research, New York.
- [32] Engle, Robert F. and Asger Lunde (2003), Trades and quotes: a bivariate process, *Journal of Financial Econometrics* 1, 2, 159–188.
- [33] Engle, Robert F. and Jeffrey R. Russell (1998), Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127–1162.
- [34] Florens, Jean-Pierre and Denis Fougere (1996), Noncausality in continuous time, *Econometrica* 64, 5, 1195–1212.
- [35] Forbes, Kristin J. and Roberto Rigobon (2002), No contagion, only interdependence: measuring stock market co-movements, *Journal of Finance* 57, 5, 2223–2261.

- [36] Geweke, John (1978), Temporal aggregation in the multiple regression model, *Econometrica* 46, 3, 643–661.
- [37] Ghysels, Eric (2012), Mixed frequency vector autoregression models, Working paper, Department of Economics, University of North Carolina at Chapel Hill.
- [38] Giesecke, Kay, 2008, Portfolio Credit Risk: Top-Down vs. Bottom-Up Approaches, In *Frontiers in Quantitative Finance: Credit Risk and Volatility Modeling*, R. Cont (Ed.), Wiley.
- [39] Glosten, Lawrence R. and Paul R. Milgrom (1985), Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.
- [40] Granger, Clive W.J. (1969), Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* 37, 3, 424–438.
- [41] Granger, Clive W.J. (1988), Some recent developments in a concept of causality, *Journal of Econometrics* 39, 199–211.
- [42] Hannan, E. J., 1967, *Time Series Analysis*, Springer.
- [43] Hasbrouck, Joel (1991), Measuring the information content of stock trades, *Journal of Finance* 46, 1, 179–207.
- [44] Hasbrouck, Joel (1999), Trading fast and slow: security market events in real time. Working Paper, Stern School of Business, New York University.
- [45] Hasbrouck, Joel (2007), *Empirical market microstructure: the institutions, economics and economics of securities trading*, Oxford University Press.
- [46] Haugh, Larry D. (1976), Checking the independence of two covariance-stationary time series: a univariate residual cross correlation approach, *Journal of American Statistical Association*, 71, 378–385.
- [47] Hautsch, Nikolaus and Luc Bauwens (2006), Stochastic Conditional Intensity Model, *Journal of Financial Econometrics*, 4, 3, 450–493.
- [48] He, Zhiguo and Wei Xiong (2012), Rollover risk and credit risk, *Journal of Finance* 67, 2, 391–429.

- [49] Heinen, Andréas and Erick Rengifo (2007), Multivariate autoregressive modeling of time series count data using copulas, *Journal of Empirical Finance*, 14, 4, 564–583.
- [50] Hiemstra, Craig and Jonathan D. Jones (1994), Testing for linear and non-linear Granger causality in the stock price-volume relation, *Journal of Finance* 49, 5, 1639–1664.
- [51] Hong, Yongmiao (1996a), Consistent testing for serial correlation of unknown form, *Econometrica* 64, 4, 837–864.
- [52] Hong, Yongmiao (1996b), Testing for independence between two covariance stationary time series, *Biometrika* 83, 3, 615–625.
- [53] Jarrow, Robert and Fan Yu (2001), Counterparty risk and the pricing of defaultable securities, *Journal of Finance* 56, 5, 1765–1799.
- [54] Jarrow, Robert A., and Stuart M. Turnbull, Pricing Derivatives on Financial Securities Subject to Credit Risk, *Journal of Finance* 50, 1995, 53–85.
- [55] Jorion, Philippe and Gaiyan Zhang (2007), Good and bad credit contagion: Evidence from credit default swaps, *Journal of Financial Economics* 84, 3, 860–883.
- [56] Jorion, Philippe and Gaiyan Zhang (2009), Credit contagion from counterparty risk, *Journal of Finance* 64, 5, 2053–2087.
- [57] Kiefer, Nicholas M., 1988, Economic Duration Data and Hazard Functions. *Journal of Economic Literature* 26, 646–679.
- [58] Kwok, Simon S., and Wai Keung Li, 2008, On Diagnostic Checking of the Autoregressive Conditional Intensity Model, *Canadian Journal of Statistics* 36, 561–576.
- [59] Lancaster, Tony, 1979, Econometric Methods for the Duration of Unemployment, *Econometrica* 47, 939–956.
- [60] Lando, David, 2004, *Credit Risk Modeling: Theory and Applications*, Princeton University Press.
- [61] Lando, David and Mads Stenbo Nielsen (2010), Correlation in corporate

defaults: Contagion or conditional independence? *Journal of Financial Intermediation* 19, 3, 355–372.

- [62] Lehmann, Erich Leo, and Joseph P. Romano, 1997, *Testing Statistical Hypotheses*, 2nd Edition, Springer.
- [63] Li, Yingying, Per A. Mykland, Eric Renault, Lan Zhang and Xinghua Zheng (2010), Realized volatility when sampling times can be endogenous, Working paper.
- [64] Marcet, Albert (1991), Temporal aggregation of economic time series. In: Hansen, L.P., Sargent, T.J. (Eds.), *Rational Expectations Econometrics*. Westview Press, Bolder, pp. 237–282.
- [65] Marcellino, Massimiliano (1999), Some consequences of temporal aggregation in empirical analysis, *Journal of Business and Economic Statistics* 17, 1, 129–136.
- [66] McCrorie, J. Roderick and Marcus J. Chambers (2006), Granger causality and the sampling of economic processes, *Journal of Econometrics* 132, 2, 311–336.
- [67] Meyer, Paul A., 1971, Démonstration Simplifiée d’un Théorème de Knight. In *Séminaire de Probabilités, V (Univ. Strasbourg, année universitaire 1969–1970)*, Lecture Notes in Mathematics 191, Springer, New York, 191–195.
- [68] Ogata, Yoshiko, 1978, The Asymptotic Behaviour of Maximum Likelihood Estimators for Stationary Point Processes, *Annals of the Institute of Statistical Mathematics* 30, 243–261.
- [69] Pacurar, Maria (2008), Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature, *Journal of economic surveys* 22, 4, 711–751.
- [70] Protter, Philip E. (2004), *Stochastic integration and differential equations*, Springer.
- [71] Ramlau-Hansen, Henrik (1983), Smoothing counting process intensities by means of kernel functions, *Annals of Statistics* 11, 2, 453–466.
- [72] Renault, Eric, Khalid Sekkat and Ariane Szafarz (1998), Testing for spurious causality in exchange rates, *Journal of Empirical Finance* 5, 1, 47–66.

- [73] Renault, Eric and Bas J.M. Werker (2010), Causality effects in return volatility measures with random times, *Journal of Econometrics* 160, 1, 272–279.
- [74] Russell, Jeffrey R. (1999), Econometric modeling of multivariate irregularly-spaced high-frequency data, Working paper, Graduate School of Business, University of Chicago.
- [75] Sims, Christopher A. (1971), Discrete approximations to continuous time distributed lags in econometrics, *Econometrica* 39, 3, 545–563.
- [76] Sims, Christopher A. (1972), Money, income, and causality, *American Economic Review* 62, 4, 540–552.
- [77] Sun, Yixiao, Peter C.B. Phillips and Sainan Jin (2008), Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica* 76, 1, 175–194.
- [78] Yu, Fan (2007), Correlated defaults in intensity-based models, *Mathematical Finance* 17, 2, 155–173.